

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:32:14 ; Search time 34.23 Seconds
(without alignments)
342.345 Million cell updates/sec

Title: US-09-466-935-2
Perfect score: 1061
Sequence: 1 MLEWWEFAYLLTSLITLSP.....KIFGSLPMLVGLIASARHA 205

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_0401:*
1: /SID2/gcgdata/geneseq/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/AA2001.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1061	100.0	205	21	E. coli L-homoseri
2	1061	100.0	205	21	E. coli RhtB prote
3	195.5	18.4	212	21	Escherichia coli y
4	155	14.6	153	17	Mel-linked mlgA ge
5	155	14.6	153	20	S. colwelliana Mlg
6	144	13.6	130	21	Human OREF ORF171
7	126.5	11.9	223	21	Escherichia coli y
8	126	11.9	195	21	Escherichia coli y
9	123.5	11.6	229	21	A. vltis hyperseus
10	122	11.5	206	21	E. coli L-threonin
11	113	10.7	211	21	Escherichia coli y

12	89.5	8.4	236	18	M37715	C. glutamicum Lys
13	80	7.5	484	20	Y04104	Arthrobacter nicot
14	80	7.5	517	20	Y04105	Arthrobacter nicot
15	79.5	7.5	307	20	Y38701	Neisseria gonorrhoe
16	79.5	7.5	307	20	Y38702	Neisseria gonorrhoe
17	79	7.4	596	20	Y09518	C. elegans dopamin
18	79	7.4	615	20	Y09517	C. elegans dopamin
19	78.5	7.4	513	20	Y38788	Neisseria gonorrhoe
20	78.5	7.4	513	20	Y38789	Neisseria gonorrhoe
21	77	7.3	496	18	M20694	H. pylori putative
22	77	7.3	539	18	M20694	H. pylori putative
23	76.5	7.2	213	21	G51077	Arabidopsis thalia
24	76.5	7.2	242	21	G51076	Arabidopsis thalia
25	76.5	7.2	243	21	G12759	Arabidopsis thalia
26	76.5	7.2	246	21	G51075	Arabidopsis thalia
27	76.5	7.2	272	21	G12758	Arabidopsis thalia
28	76.5	7.2	276	21	G12757	Arabidopsis thalia
29	76	7.2	553	20	Y01651	Arabidopsis thalia
30	76	7.2	553	21	B20579	A protein with cat
31	75.5	7.1	277	20	Y38699	Mouse OCTN1 amino
32	75.5	7.1	396	17	W02112	Neisseria meningit
33	75.5	7.1	396	17	W02113	Human homologue of
34	74.5	7.0	170	20	Y29197	Murine int6 protei
35	74.5	7.0	465	20	Y35106	Amino acid sequenc
36	74	7.0	277	20	Y38698	Chlamydia pneumoni
37	74	7.0	551	20	Y01649	Neisseria meningit
38	74	7.0	783	21	Y96164	A protein with cat
39	73.5	6.9	153	19	W80604	S. typhimurium pox
40	73.5	6.9	375	18	W20731	Eucalyptus grandis
41	73.5	6.9	955	21	B16338	Cellulose synthase
42	73.5	6.9	974	19	W73308	Cotton cellulose s
43	73	6.9	974	19	W60039	Rodent GPR2 protei
44	73	6.9	362	21	B01449	Amino acid sequenc
45	72	6.8	468	20	Y34877	

ALIGNMENTS

RESULT 1	
ID	Y99597 standard; Protein; 205 AA.
Y99597	
XX	
AC	Y99597:
XX	
DT	08-SEP-2000 (first entry)
XX	
DE	E. coli L-homoserine resistance protein, RhtB.
XX	
KW	L-homoserine resistance; L-homoserine synthesis; rhtB;
KW	L-threonine; L-valine; L-leucine.
XX	
OS	Escherichia coli.
XX	
PN	EP1013765-A1.
XX	
PD	28-JUN-2000.
XX	
PF	20-DEC-1999; 99EP-0125406.
XX	
PR	23-DEC-1998; 98RU-0123511.
XX	
PA	(AJIN) AJINOMOTO KK.
XX	
PI	Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokmakova IL;
XX	
DR	WPI: 2000-414602/36.
XX	
DR	N-PSDB; A48442.
XX	
PT	Novel Escherichia bacterium having enhanced L-threonine resistance due
PT	to enhanced RhtC protein activity, used to produce L-threonine,
PT	L-homoserine, L-valine and L-leucine -
XX	

PS Claim 2; Page 13-14; 24pp; English.

XX The present sequence is the L-homoserine resistance protein, RhtB, from
CC Escherichia coli. The coding sequence may be used to impart L-homoserine
CC resistance on E. coli bacteria, which would be useful for producing
CC a high yield of L-homoserine. L-homoserine resistance means that the
CC bacteria will be able to grow on a minimal medium containing
CC L-homoserine at a concentration at which the corresponding wild-type
CC strain would not grow. Since the transformed bacteria can grow on the
CC minimal medium, it can synthesize L-homoserine, which accumulates. The
CC accumulated amino acids can then be removed from the culture medium.
CC The bacterium of the present invention may also be used to synthesize
CC L-threonine, L-valine and L-leucine at increased levels.

XX Sequence 205 AA;

XX SQ

Query Match 100.0%; Score 1061; DB 21; Length 205;
Best Local Similarity 100.0%; Pred. No. 3.9e-113;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLEWMFAVLLTSLITLSPSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGRGV 60
DB 1 mLEWwfaYllTsIlLtLspsgaIntmttslnhgyPaGgyvCwaSdrGdsyCaGrGv 60
QY 61 GILFSRSVIAFEYKWKAGAAYLIMLGIOQWRAAGATDKSLASTOSRRHLFORAVVNL 120
DB 61 gILfSrsvIAfeYlKwGaAaYlIwIgIqWraaGaIdKslAsTosRRhlfOraVvNl 120
QY 121 NPKSIVFLAALFPQFTMPQOPOLMOYIVLGVTIVVDIIVMIGYATLAORIALMIGPKQ 180
DB 121 nPkSivfLaAlfPqfTmPqOpOlMoYiVlGvTtIvVdIiVmIgYaTlAoRiAlMiGpKq 180
QY 181 MKALNKIFGSLFMLVGALLASARHA 205
DB 181 mKaLnKIfgSlfMlVgAlLaSaRha 205

RESULT 2

Y79298
ID Y79298 standard; Protein; 205 AA.

AC Y79298;

DT 18-JUL-2000 (first entry)

DE E. coli RhtB protein imparting homoserine resistance.

KW Homoserine resistance; RhtB protein; L-homoserine; L-alanine;

KM L-isoleucine; L-valine; L-threonine.

OS Escherichia coli.

PN EP994190-A2.

PD 19-APR-2000.

PF 20-SEP-1999; 99EP-0118581.

PR 13-OCT-1998; 98RU-0118425.

PA (AJIN) AJINOMOTO CO INC.

PI Liivshits VA, Zakataeva NP, Aleoshin VV, Belareova AV;

DR N-PSDB; Z94405.

PT Novel RhtB protein, useful for generation of L-homoserine resistance in

PT Escherichia bacteria and large-scale production of e.g. L-homoserine

PT and L-alanine

XX

PS Claim 1; Page 11-12; 14pp; English.

XX The present sequence is that of the novel Escherichia coli K-12
CC protein, RhtB, which participates in resistance to L-homoserine.
CC Amplification of the rhtB gene (see Z94405) results in an
CC improvement of the amino acid productivity of E. coli. The
CC invention provides: a protein comprising the present sequence or
CC including a deletion, substitution, insertion and/or addition of 1
CC or more amino acids and having the activity of making a bacterium
CC resistant to L-homoserine; DNA encoding RhtB; a bacterium,
CC especially of the genus Escherichia, in which L-homoserine
CC resistance is enhanced by amplifying the copy number or increasing
CC the expression rate of the rhtB DNA, the DNA being carried on a
CC multicopy vector or on a transposon; and a method for producing an
CC amino acid by cultivating the bacterium in a culture medium to
CC produce and accumulate the amino acid in the medium, from which it
CC is recovered. The method is used for the production of
CC L-homoserine, L-alanine, L-isoleucine, L-valine or L-threonine (all
CC claimed).

XX Sequence 205 AA;

XX SQ

Query Match 100.0%; Score 1061; DB 21; Length 205;
Best Local Similarity 100.0%; Pred. No. 3.9e-113;
Matches 205; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MLEWMFAVLLTSLITLSPSGAINTMTTSLNHGYPAGGVYCWASDRGDSYCAGRGV 60
DB 1 mLEWwfaYllTsIlLtLspsgaIntmttslnhgyPaGgyvCwaSdrGdsyCaGrGv 60
QY 61 GILFSRSVIAFEYKWKAGAAYLIMLGIOQWRAAGATDKSLASTOSRRHLFORAVVNL 120
DB 61 gILfSrsvIAfeYlKwGaAaYlIwIgIqWraaGaIdKslAsTosRRhlfOraVvNl 120
QY 121 NPKSIVFLAALFPQFTMPQOPOLMOYIVLGVTIVVDIIVMIGYATLAORIALMIGPKQ 180
DB 121 nPkSivfLaAlfPqfTmPqOpOlMoYiVlGvTtIvVdIiVmIgYaTlAoRiAlMiGpKq 180
QY 181 MKALNKIFGSLFMLVGALLASARHA 205
DB 181 mKaLnKIfgSlfMlVgAlLaSaRha 205

RESULT 3

B01787
ID B01787 standard; Protein; 212 AA.

AC B01787;

DT 03-JAN-2001 (first entry)

DE Escherichia coli Yeas amino acid excretion protein.

KW E. coli; yeas gene; amino acid production; excretion protein gene;

KM amino acid excretion protein.

OS Escherichia coli.

PN EP1016710-A2.

PD 05-JUL-2000.

PF 17-DEC-1999; 99EP-0125263.

PR 30-DEC-1998; 98RU-0124016.

PR 09-MAR-1999; 99RU-0104431.

PA (AJIN) AJINOMOTO CO INC.

PI Liivshits VA, Zakataeva NP, Nakanishi K, Aleoshin VV, Troshin PV;

PI Tokmakova IL;

XX

DR	WP1: 2000-414802/36.
DR	N-PSDB; A52689.
XX	
PT	Increased production of L-amino acids by an Escherichia bacterium
XX	comprises increasing the expression amount of an L-amino acid excretion
PT	protein -
XX	
PS	Claim 1; Page 21; 29pp; English.
XX	
CC	The present sequence is the Yeas amino acid excretion protein from
CC	Escherichia coli. This protein is involved in the production of amino
CC	acids, and an increase in its expression leads to an increased
CC	accumulation of amino acids in the cell. In this case, an increase in
CC	lysine, alanine, valine, histidine, isoleucine, glutamic acid and proline
CC	is achieved if multiple copies of its gene are transfected into a
CC	bacterium. The bacterium used is E. coli.
XX	
SO	Sequence 212 AA;
XX	
Query Match	18.4%; Score 195.5; DB 21; Length 212;
Best Local Similarity	28.6%; Pred. No.2.4e-14;
Matches	61; Conservative 45; Mismatches 82; Indels 25; Gaps 10.
QY	3 LEWFAVLNLSILTLPSGAI---NTMTSLNHGYPAG-GVYCWASDRGTGDSYC--A 55
DB	1 lny-tylgafifvlypnpnlfvlnksvssgmkgyllaacyf-----igdavlml 60
QY	56 GMRVGLFRSRYIAEVLKMGAAATLMLGIQWRRAAGALDK---SLASTOSRRH--L 110
DB	61 awagavaliiktpilfnhivrylgafilylgskilyat---lkgknskaaksdepygga 116
QY	111 FORAVEFNLTNPKSIVFELALFPQFIMPOQPOL-MQYIVLGVTTIVDIYIMGYATIAQ 169
DB	117 fkrallstlnpkaillfyvsffvgfidrnaphgtglsffilaatlslvscyifslilissa 176
QY	170 RIALWIKGRKM-KALNKITFGSEFMVIGALLAS 201
DB	177 fvtgyirtkkklakvgnslgimfvgyfaariat 209
XX	
RESULT 4	
ID	R87527
XX	R87527 standard; Protein: 153 AA.
AC	R87527;
XX	
DT	01-JUL-1996 (first entry)
XX	
DE	Me1-linked mlga gene product.
XX	
KW	Marine mel1a; selectable marker; oyster larva settlement;
KW	plimentation; melanin; p-hydroxyphenylpyruvate-hydroxylase;
XX	mlga gene.
XX	
OS	Shewanella colwelliana strain LSTDYF.
XX	
PN	US5474933-A.
XX	
PD	12-DEC-1995.
XX	
PF	21-MAR-1990; 90US-0496804.
XX	
PR	08-NOV-1993; 93US-0148945.
PR	21-MAR-1990; 90US-0496804.
PR	10-NOV-1992; 92US-0974837.
XX	
PA	(UYMA-) UNIV MARYLAND BALTIMORE.
XX	
PI	Fuqua WC, Weiner RM;
XX	
DR	WP1: 1996-039515/04.
DR	N-PSDB; T06767.

[illegible]

CC related to the Mela protein of the invention. The invention relates to
CC Mela proteins from marine bacterium of the genus *Shewanella*, *Vibrio* or
CC *Hyphomonas*. The Mela protein is active in catalysing production of
CC melanin synthesis, as well as components of marine exopolysaccharides.
CC The Mela protein can be used as a uv blocker in sunscreen, pigment and
CC dye compositions useful in cosmetic applications, as tints for glass or
CC in paints and coatings for plastics, synthetic resins and fabrics, rubber
CC and wood.

CC Sequence 153 AA:

Query Match 14.6%; Score 155; DB 20; Length 153;
Best Local Similarity 29.7%; Pred. No. 6.6e-10;

Matches 38; Conservative 21; Mismatches 51; Indels 18; Gaps 2;

QY 57 MRGVGTLFRSYIA-----FEVLMKAGAYLIMLGIOQWRAAGAI-DL 98

DB 18 MIMVGELAGVALVALAAGVASMMLNYPQLFDLKWVGGLYGLYGLISMWRKGMANL 77

QY 99 KSLASTQSRRLHFORAVFVNLTNPKSIVFLAALFPQFTMPQOQPLMQYIVLGVTTIVDI 158

DB 78 dntsqisrnatlgtgfvfalaanpkywalmisllppflsvdgaiaqqlmvlisimtlef 137

QY 159 IVMIGYAT 166

DB 138 fsmLAYas 145

RESULT 6

BA0407

ID B40407 standard; Protein: 130 AA.

AC B40407;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF171 polypeptide sequence SEQ ID NO:342.

XX Human; open reading frame; ORFX; detection; cytosolic; hepatotropic;
XX vulnary; antiparasitic; antiparkinsonian; neurotrophic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiac;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antipneumatic; antihypertoid;
XX antineutemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; cartilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive.

XX Homo sapiens.

XX W0200058473-A2.

XX 05-OCT-2000.

XX 31-MAR-2000; 2000MO-US08621.

XX 31-MAR-1999; 99US-0127607.

XX 02-APR-1999; 99US-0127636.

XX 03-APR-1999; 99US-0127728.

XX 30-MAR-2000; 2000US-0540763.

XX (CURA-) CURAGEN CORP.

XX Shimkets RA, Leach M;

XX WPI; 2000-602362/57.

XX N-PSDB; C74616.

XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
PS Claim 11; Page 608; 5507pp; English.

CC C74446 to C77606 encode the proteins given in B40237 to B43397, which

CC represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytosolic; hepatotropic; vulnary;
CC antiparasitic; antiparkinsonian; neurotrophic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiac; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antineumatic; antihypertoid; and antipneumatic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 130 AA:

Query Match 13.6%; Score 144; DB 21; Length 130;

Best Local Similarity 31.0%; Pred. No. 9.5e-09;

Matches 40; Conservative 22; Mismatches 61; Indels 6; Gaps 4;

QY 79 AAYLIMLGIOQWRAAGALDKSLASTQSRRLHFORAVFVNLTNPKSIVFLAALFPQFTMP 138

DB 3 aayllylayatwtdrsafamndpvtatarslllrgflnlhnpkltiflalfpqfvtp 62

QY 139 --QQPOLMQYIVLGVTTIVDIIMIGYATLAQ--RTALMIKGRKMKALNKITFGSLFML 194

DB 63 ggtupal-qmlvlsygvmtamtlavfvygllanvfrav-vesprgvnwlrfsafafag 120

QY 195 VGALLASAR 203

DB 121 lglhlafeq 129

RESULT 7

B01786

ID B01786 standard; Protein: 223 AA.

AC B01786;

XX 03-JAN-2001 (first entry)

XX *Escherichia coli* yahn amino acid excretion protein.

XX E. coli; yahn gene; amino acid production; excretion protein gene;

XX amino acid excretion protein.

XX *Escherichia coli*.

XX EP1016710-A2.

XX 05-JUL-2000.

XX 17-DEC-1999; 99EP-0125263.

XX 30-DEC-1998; 98RU-0124016.

XX 09-MAR-1999; 99RU-0104431.

XX (AJIN) AJINOMOTO CO INC.

XX	30-DEC-1998;	96RU-0124016.
PR	09-MAR-1999;	99RU-0104431.
XX		
PA	(AJIN)	AJINOMOTO CO INC.
XX		
PI	Iivshits Vb.,	Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
PI	Tokhmakova IL;	
DR	WPI, 2000-414802/36.	
DR	N-PSDB; A52691.	
XX		
PT	Increased production of L-amino acids by an Escherichia bacterium	
PT	complies increasing the expression amount of an L-amino acid excretion	
PT	protein -	
XX		
PS	Claim 1; Page 25; 29pp; English.	
XX		
CC	The present sequence is the <i>ygfa</i> amino acid excretion protein from	
CC	<i>Escherichia coli</i> . This protein is involved in the production of amino	
CC	acids, and an increase in its expression leads to an increased	
CC	accumulation of amino acids in the cell. In this case, an increase in	
CC	arginine, glutamic acid and lysine is achieved if multiple copies of the	
CC	gene are transfected into a bacterium. The bacterium used is <i>E. coli</i> .	
XX		
XX	Sequence	211 AA;

```

Query Match=      10.7%; Score 113; DB 21; Length 211;
Best Local Similarity 24.9%; Pred. No. 6,2e-05;
Matches    52; Conservative   36; Mismatches   95; Indels    26; Gaps     10;

QY          7 FAV-----LILSLIILTSPGSAIMWTMTSINNGYPAG-GVYCMAISDPTGSVCAGARG 59
              |::||| ||| : ::||| |::||| |
DB           2 fsyyfgagialgaamllplpgnafv--mmgqirryhmatalcalcaidsdv--ljcaglfg 57
              |::||| |::||| |::||| |::||| |
QY          60 VGLTSRSVIATFEVLKMGAGAAYITWLGTIOQMRAAQAIDLSIAST*-QSRRHLFQRAY 115
              |::||| |::||| |::||| |::||| |
DB           58 gsallmgpswllaivtvgvaflilwygfafkxamsnhe-lasaevmkgwtklialmll 116
              |::||| |::||| |::||| |::||| |
QY          116 EVNLTNP---KSIVFLALFPDFIMPDPOLMOXYIVLGVTTVVDIIVMIGYAHTAQRI 171
               | | |
DB           117 awvlrmphvyldtfvvlgslsgd--ldvepk--rwfalg--tlssasflwfjglailaaawl 170
                |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY          172 ALMIKGPKOMALKIKFGSLFMWGAULTA 200
                 |::|::|::|::|::|::|::|
DB           171 aplrlrakagrlnlvgcwmvfiaqlqa 199
                  |::|::|::|::|::|::|::|

RESULT      12
W37715
XX ID W37715 standard; Protein; 236 AA.
AC
GC W37715;
TC
DT 12-MAR-1998 (first entry)
DE
EE C.: glutamicum Lys E protein (lysine export protein).
FX XX
KM lysG; lysE; ORF3; lysine transport; regulatory protein; export protein;
KW Microbial production; amino acid; animal feed additive.
XX
OS Corynebacterium glutamicum.
XX
PN DE19548222-A1.
XX PD
XX 26-JUN-1997.
XX PE
PF 22-DEC-1995; 95DE-1048222.
PR 22-DEC-1995; 95DE-1048222.
PX
PA (KERJ ) FORSCHUNGSENZENTRUM JUELICH GMBH.
```

XX	Eggeling L, Sahm H, Wille M:
PI	
XX	WPI, 1997-333867/31.
DR	N-PSDB; T96816.
DR	
XX	
PT	Increasing microbial production of amino acids, especially lysine -
PT	by improving export carrier activity or corresponding gene
PT	expression, also new export and regulatory genes from
PT	Corynebacterium
XX	
PS	Claim 42; Page 10; 16pp; German.
XX	
CC	This sequence is the LysE protein product, a lysine export protein.
CC	LysE and LysE encode a lysine transport regulatory protein and an export
CC	protein, respectively. Microbial production of amino acids (A) is
CC	improved by increasing the export-carrier activity and/or the export gene
CC	expression in a microorganism that produces (A). The method is
CC	specifically used to increase production of lysine, used as an animal
CC	feed additive. Other (A) are variously useful as pharmaceuticals,
CC	condiments and intermediates for fine chemicals. This method increases
CC	the amount of (A) secreted into the culture medium. Export of (A) has
XX	been found to depend on a single gene.
XX	
Sequence	236 AA;
50	

	Query Match Similarity	8.4%;	Score 89.5;	DB 18;	Length 236;	
	Best Local Similarity	20.6%;	Pred. No. 0.035;			
	Matches	36;	Conservative	25;	Mismatches	79; Indels 35; Gaps
Qy	3	LEWHFVYLITSLITLSPSGAINTMTISL-NHGYPRAGYVCMAASRDTSICACMRGCG	61			
Dd	4	mellitgtlllgasllisgpnaylvkgyikregllaavllvcllsdvf--lilaqlgv	61			
Qy	62	TIFSRSVAFEVLWAGAAAYLIWLGIQCWRAGAIDLKSLASTQSRRHLFORAFVNLTN	121			
Dd	62	llsnaaaplvlldlmwgglayllwfawmaakda-----mtn	96			
Qy	122	KPSIVLAALPFQFMPOCPOLMQYVLCGVTVVDIIMVGATFLAKRIALMIK	176			
Dd	97	kvea-----pqllteetptypddtprigsavaldtnrvrevsvdkgrvwk	144			
RESULT	13					
Y04104	ID	Y04104 standard; Protein; 484 AA.				
XX	AC	Y04104;				
XX	DE	10-JUN-1999 (first entry)				
XX	DT	Arthrobacter nicotlinovorans levan fructotransferase protein #1.				
XX	FW	Arthrobacter nicotlinovorans, levan fructotransferase.				
OS	Arthrobacter nicotlinovorans.					
XX	JPI1069978-A.					
FN	16-MAR-1999.					
PD	28-AUG-1997;	97JP-0232421.				
XX	28-AUG-1997;	97JP-0232421.				
XX	(NIOC) NIPPON OIL CO LTD.					
PA	WPI: 1999-247463/21.					
DR	N-Psdb; X19826.					
XX	Levan fructotransferase gene - for recombinant production of levan					
XX	fructotransferase					
PT						

XX Claim 1; Page 5-8; 14pp; Japanese.

CC The present sequence represents *Arthrobacter nicotinovorus* levan
fructotransferase. The present invention also describes a method
CC for the preparation of levan fructotransferase in which a transformant
CC is cultured in a medium and levan fructotransferase is collected from
CC the culture. The method can prepare levan fructotransferase in a
CC large amount.

XX Sequence 484 AA;

Query Match 7.5%; Score 80; DB 20; Length 484;
Best Local Similarity 19.3%; Pred. No. 1.1;
Matches 41; Conservative 34; Mismatches 81; Indels 56; Gaps 7;

OY 3 LEW---WFA-----YLLTSIIITLSPSGAINTMTSLNHG 35
DB 254 ldkwcdyaavtvpaveapetkrlataamnnkyaarnvpdaadgynqnsitrelrlle 313
OY 36 YPAGGVTC-----WASDRTGDSYCAGMKGVTLSRSYIAFEVILKMAAGAAVLIW 84
DB 314 rqsqgwyllstcpypalsnyatset-----llpdrtrngsfvlpwsgrayele 361
OY 85 LGIQOMRAGAIDKSLASTOSRHL---FORAVFVNLTPKSIYVFLAALFPQFIIMPQO 140
DB 362 ldl-swtdaanvsgvrsdgsrhtnlgkygdelydrasseqsgyalapylraapid 420
OY 141 POLMQ-YIVLGVTITVVDIIMIGVATLAORI 171
DB 421 anarsvhlrlftvdtsqevfvtvsghtvsgqv 452

RESULT 14
Y04105
ID Y04105 standard; Protein: 517 AA.

XX Y04105;

DT 10-JUN-1999 (first entry)

DE *Arthrobacter nicotinovorus* levan fructotransferase protein #2.

KW *Arthrobacter nicotinovorus*; levan fructotransferase.

OS *Arthrobacter nicotinovorus*.

XX JP1069978-A.

PD 16-MAR-1999.

PF 28-AUG-1997; 97JP-0232421.

PR 28-AUG-1997; 97JP-0232421.

PA (NIOC) NIPPON OIL CO LTD.

DR WPI: 1999-247463/21.

XX N-PSDB: X19827.

PT Levan fructotransferase gene - for recombinant production of levan

XX fructotransferase

XX Claim 3; Page 8-9; 14pp; Japanese.

CC The present sequence represents *Arthrobacter nicotinovorus* levan
fructotransferase. The present invention also describes a method
CC for the preparation of levan fructotransferase in which a transformant
CC is cultured in a medium and levan fructotransferase is collected from
CC the culture. The method can prepare levan fructotransferase in a
CC large amount.

SQ Sequence 517 AA;

Query Match 7.5%; Score 80; DB 20; Length 517;
Best Local Similarity 19.3%; Pred. No. 1.2;
Matches 41; Conservative 34; Mismatches 81; Indels 56; Gaps 7;

OY 3 LEW---WFA-----YLLTSIIITLSPSGAINTMTSLNHG 35
DB 287 ldkwcdyaavtvpaveapetkrlataamnnkyaarnvpdaadgynqnsitrelrlle 346
OY 36 YPAGGVTC-----WASDRTGDSYCAGMKGVTLSRSYIAFEVILKMAAGAAVLIW 84
DB 347 rqsqgwyllstcpypalsnyatset-----llpdrtrngsfvlpwsgrayele 394
OY 85 LGIQOMRAGAIDKSLASTOSRHL---FORAVFVNLTPKSIYVFLAALFPQFIIMPQO 140
DB 395 ldl-swtdaanvsgvrsdgsrhtnlgkygdelydrasseqsgyalapylraapid 453
OY 141 POLMQ-YIVLGVTITVVDIIMIGVATLAORI 171
DB 454 anarsvhlrlftvdtsqevfvtvsghtvsgqv 485

RESULT 15
Y38701
ID Y38701 standard; Protein: 307 AA.

XX Y38701;

DT 08-OCT-1999 (first entry)

DE *Neisseria gonorrhoeae* antigen encoded by partial ORF104.

KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;
treatment; *Neisseria* infection; meningitis; septicemia; gonorrhea.

OS *Neisseria gonorrhoeae*.

XX WO9924578-A2.

PN 20-MAY-1999.

PF 09-OCT-1998; 98WC-IB01665.

PR 01-SEP-1998; 98GB-0019016.

PR 06-NOV-1997; 97GB-0023516.

PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.

PR 27-NOV-1997; 97GB-0025158.

PR 10-DEC-1997; 97GB-0026147.

PR 14-JAN-1998; 98GB-0000759.

XX (CHIR-) CHIRON SPA.

PA Grandi G, Maignani V, Piazza M, Rappuoli R, Scarlato V;

XX WPI: 1999-327407/27.

DR proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for

XX diagnosis, treatment and prevention of infection

XX Claim 4; Page 252; 524pp; English.

CC Amino acid sequences Y38499-Y38944 represent *Neisseria meningitidis*
and *N. gonorrhoeae* antigenic proteins. They are encoded by open
CC reading frames (ORFs) 211972-212358. The antigenic proteins
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*
CC infections, such as meningitis, septicemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.

us-09-466-935-2.rag

Page 9

Query Match	7.5%;	Score 79.5;	DB 20;	Length 307;
Best Local Similarity	27.0%;	Pred. No. 0.68;		
Matches	41;	Conservative	25;	Mismatches 69;
				Indels 17;
				Gaps 6;

```

Oy 61 GTELSKRVIAFEVLEKKAAGAALVIMGTIOQWRAAGI--DLKSLASIPGSRKHLEQ----RA 114
    ||| : ||| : ||| : ||| :
Db 21 gtl---plavqgltkvdapltlwv---rfvtaavilfvllalgyrlpkrrdswshfsl 74
    : : : : :
Oy 115 VFVNLINPKRIVPLALPQGITMPQOPOLMOGYIVGTTIVDITIVMGVATLAORIALM 174
    : : : : :
Db 75 lllygvtsianfvlisqglhyispttqyl-wqslpfmllvwyglvltkdmteaqaqklylv 133
    : : : : :
Oy 175 IKSGPKOKALNKLIFGSLFMV---WGALIASA 202
    : : | | | | |
Db 134 lllyglimfndkfgeislgslyakgavlllcaa 165

```

Search completed: May 6, 2001, 14:38:48
Job time: 394 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:32:29 ; Search time 21.59 Seconds
(without alignments)
182,410 Million cell updates/sec

Title: US-09-466-935-2
Perfect score: 1061
Sequence: 1 MTELEMPAYLITLITLSP.....KIFGSLFMVGLASARHA 205

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 185757 seqs, 19210857 residues

Total number of hits satisfying chosen parameters: 185757

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued Patents, AA: *
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep.*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep.*
3: /cgn2_6/ptodata/2/1aa/6A.COMB.pep.*
4: /cgn2_6/ptodata/2/1aa/6B.COMB.pep.*
5: /cgn2_6/ptodata/2/1aa/PCrUS.COMB.pep.*
6: /cgn2_6/ptodata/2/1aa/Backfile1.pep.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query Match	Length	DB ID	Description
1	155	14.6	153	2	US-08-476-254-7
2	155	14.6	153	6	5474933-4
3	72	6.8	503	4	US-09-068-195-24
4	70	6.6	428	6	5432081-9
5	68.5	6.5	534	4	US-08-875-223-8
6	68.5	6.5	857	2	US-08-448-603A-30
7	68.5	6.5	857	3	US-09-134-075-30
8	68	6.4	456	6	5432081-7
9	67.5	6.4	1302	1	US-08-232-537-2
10	66.5	6.3	291	1	US-08-358-171-2
11	66.5	6.3	291	2	US-09-090-947-2
12	66.5	6.3	344	2	US-08-468-812-2
13	66.5	6.3	898	3	US-08-863-102-1
14	66.5	6.3	898	3	US-08-863-102-4
15	66	6.2	772	1	US-08-802-141-4
16	66	6.2	1098	3	US-08-726-214-10
17	65.5	6.2	379	1	US-08-227-108-18
18	65.5	6.2	379	2	US-09-073-674-18
19	64.5	6.1	380	2	US-08-472-659-34
20	64.5	6.1	380	2	US-08-474-661-34
21	64.5	6.1	453	2	US-08-611-977-34
22	64.5	6.1	453	2	US-08-244-205-13
23	64.5	6.1	453	5	PCr-US92-10284-13
24	64.5	6.1	489	2	US-08-752-307B-3
25	64.5	6.1	530	2	US-08-752-307B-2
26	64.5	6.1	642	1	US-08-706-936-3
27	64	6.0	1052	2	US-08-852-806-2

28	64	6.0	1052	4	US-09-163-669-2	Sequence 2, Appl
29	63.5	6.0	426	6	5268463-8	Patent No. 5268463
30	63.5	6.0	506	3	US-08-867-352-21	Sequence 21, Appl
31	63	5.9	371	4	US-09-043-937A-4	Sequence 4, Appl
32	63	5.9	533	1	US-08-368-071-10	Sequence 10, Appl
33	63	5.9	533	1	US-08-458-181-10	Sequence 10, Appl
34	63	5.9	533	5	PCr-US93-02172-10	Sequence 10, Appl
35	63	5.9	607	1	US-07-959-943-7	Sequence 7, Appl
36	63	5.9	607	1	US-07-879-617A-12	Sequence 12, Appl
37	63	5.9	607	1	US-08-753-985-12	Sequence 12, Appl
38	63	5.9	630	1	US-07-959-943-9	Sequence 9, Appl
39	63	5.9	642	1	US-08-706-936-2	Sequence 2, Appl
40	63	5.9	653	1	US-07-782-298-2	Sequence 2, Appl
41	63	5.9	686	3	US-09-306-922-2	Sequence 2, Appl
42	63	5.9	759	2	US-08-637-759B-89	Sequence 89, Appl
43	63	5.9	759	3	US-08-871-355A-89	Sequence 89, Appl
44	62.5	5.9	197	4	US-09-402-668-2	Sequence 2, Appl
45	62.5	5.9	280	2	US-08-700-013B-15	Sequence 15, Appl

ALIGNMENTS

RESULT 1
US-08-476-254-7
; Sequence 7, Application US/08476254
; Patent No. 5846531
; GENERAL INFORMATION:
; APPLICANT: WEINER, RONALD M.
; TITLE OF INVENTION: MARINE MELA GENE
; NUMBER OF SEQUENCES: 11
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: WATSON COLE STEVENS DAVIS, P.L.L.C.
; STREET: 1400 K STREET NW
; CITY: WASHINGTON
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-2477
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/476,254
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: PODIOS III, JAMES A.
; REGISTRATION NUMBER: 31,714
; REFERENCE/DOCKET NUMBER: JAP30319C
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202 628-0088
; TELEFAX: 202 628-8034
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 153 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-476-254-7

Query Match 14.6%; Score 155; DB 2; Length 153;
Best Local Similarity 29.7%; Pred. No. 1.2e-10;
Matches 38; Conservative 21; Mismatches 51; Indels 18; Gaps 2;

QY 57 WRGCTLSRSVIA-----FEVTKAGAAVTLIMGIGQWRAGAI-DL 98
DB 18 WMWGELGALVALIAAIVGVASMLNYPOLFIDILIMVGGLYIGTIGISMIRAKGKMANL 77
QY 99 KSLASTOSRRHLFORAVFVNLTNPKSIVFLAALFPOFIMPQOPOLMQYIVLGVTIVVDI 158

Db 78 DNTSSQSNKALITGCFVTAIANPKGMAMISLLPFTISVQALAPQMLVLLSIIMTEF 137
QY 159 IWMIGYAT 166
Db 138 ESMLAYAS 145

RESULT 2
5474933-4
; Patent No. 5474933
; APPLICANT: EINER, RONALD M.; FOUOA JR., WILLIAM C.
; TITLE OF INVENTION: MARINE MEIA GENE
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,945
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 974,837
; FILING DATE: 10-NOV-1992
; APPLICATION NUMBER: 496,804
; FILING DATE: 21-MAR-1990
; SEQ ID NO: 4:
; LENGTH: 153
5474933-4

Query Match 14.6%; Score 155; DB 6; Length 153;
Best Local Similarity 29.7%; Pred. No. 1.2e-10;
Matches 38; Conservative 21; Mismatches 51; Indels 18; Gaps 2;

QY 57 WRVGVGTFRSVIA-----FEVYKMGAAVYLMIGIQWRAGAL-DL 98
Db 18 WAAVAGELAAVATAAAGVASMALNPQLPDLTKWGLYLGISMRAGKMANL 77
QY 99 KSLASTQSRRLHFORAVFNLTNPKSIYFLAALFPQFIMPQPOLMOTIVLGVTIVVDI 158
Db 78 DNTSSQSNKALITGCFVTAIANPKGMAMISLLPFTISVQALAPQMLVLLSIIMTEF 137
QY 159 IWMIGYAT 166
Db 138 ESMLAYAS 145

RESULT 3
US-09-068-195-24
; Sequence 24, Application US/09068195B
; Patent No. 6140078
; GENERAL INFORMATION:
; APPLICANT: Sanders, Jan W.
; APPLICANT: Ledebner, Adriaan M.
; APPLICANT: Venema, Gerard
; TITLE OF INVENTION: Salt-inducible Promoter Derivable from a Lactic Acid
; TITLE OF INVENTION: Bacterium, and Its Use in a Lactic Acid Bacterium for
; TITLE OF INVENTION: Production of Desired Protein
; FILE REFERENCE: Sanders-60113/0252227
; CURRENT APPLICATION NUMBER: US/09/068,195B
; EARLIER FILING DATE: 1998-07-29
; EARLIER APPLICATION NUMBER: PCT/EP97/04755
; EARLIER FILING DATE: 1997-08-20
; EARLIER APPLICATION NUMBER: EP 97200744/7
; EARLIER FILING DATE: 1997-03-13
; EARLIER APPLICATION NUMBER: EP 96202444/4
; EARLIER FILING DATE: 1996-09-05
; NUMBER OF SEQ ID NOS: 25
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 24
; LENGTH: 503
; TYPE: PRT
; ORGANISM: L. lactis MG1363
US-09-068-195-24

Query Match 6.8%; Score 72; DB 4; Length 503;
Best Local Similarity 21.1%; Pred. No. 3.6;
Matches 41; Conservative 25; Mismatches 66; Indels 62; Gaps 10;

QY 9 YLITSLITLSPGSGAINTMTSLNHGYPAGVYCMASDRTDGSCAGMRGVTLSR-- 66
Db 38 FLLGLGLMFLPALCAEMMATV--EGHKNQGIYSWVSQILGERF-----GFAAIFPQWF 90
QY 67 -----SVIAFEVL-----KWAGAAVYLMIGIQWRAGALDKSLA 102
Db 91 QTVGVVMTIYFLLGALSYVNFQALNTDPLIKFGLILIFW-----GLTESQLG 140
QY 103 STQSRRLHFORAVFNLTNPKSIYFLAALFPQFIMPQPOLMOTIVLGVTIVVDI 150
Db 141 GQGRAKLVKAGFVYGIYIPSVILFGLAAAL--PIGPNPIEIPINSHAFVDPDSQ----- 193
QY 151 VTTIVVDIIMIGY 164
Db 194 VSTLVVFSFILAY 207

RESULT 4
5432081-9
; Patent No. 5432081
; APPLICANT: JEFFERSON, RICHARD A.
; TITLE OF INVENTION: HOST CELLS TRANSFORMED WITH THE E. COLI
; GLUCONONIDE PERMEASE GENE
; NUMBER OF SEQUENCES: 10
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/138,546
; FILING DATE: 15-OCT-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 447,976
; FILING DATE: 08-DEC-1989
; APPLICATION NUMBER: 264,586
; FILING DATE: 31-OCT-1988
; APPLICATION NUMBER: 119,102
; FILING DATE: 10-NOV-1987
; SEQ ID NO: 9:
; LENGTH: 428
5432081-9

Query Match 6.8%; Score 70; DB 6; Length 428;
Best Local Similarity 26.1%; Pred. No. 4.9;
Matches 58; Conservative 27; Mismatches 101; Indels 36; Gaps 13;

QY 7 FAYLITSLITLSPGSGAINTMTSLNHGYPAGVYCMASD--RTGDSYCA-GWR----- 58
Db 14 FAFAGALFL-LSYTYDVAGVGAARAAHAYTAGIRCLRRLRCLRTSGGCEYRWGKFRPF 72
QY 59 ---GVGTLSRSVIAFEVL--KWAGA-----AYLTMIGIQWRAGALDKSLAST----- 104
Db 73 LLEGTAPIMTISVLFVMTLDMSHGSKVYAYLYLMGLGCLSYLVNIPGSLATANTQOP 132
QY 105 QSRRLHFORAVFNLTNPKSIYFLAALFPQFIMPQPOLMOTIVLGVTIVVDIIMIG 163
Db 133 QSRRLHGAARIAASLT---FVCLAFLLGSPISKNSSPEEM-VSVYHFXTYVLAIAAGVNL 187
QY 164 Y----ATLAQRIALIMIKPKOMKLINKIFS--LEML-VGAL 198
Db 168 YFICKKSTRENVAIVVAPSLNISLQTLKNNRPLFMCLGAL 229

RESULT 5
US-08-875-223-8
; Sequence 8, Application US/08875223
; Patent No. 6127175
; GENERAL INFORMATION:
; APPLICANT: VIGNE, Emmanuelle
; APPLICANT: PERRICAUDDET, Michel
; APPLICANT: DEDIEU, Jean-Francois

Mon May 7 11:08:02 2001

us-09-466-935-2.rai

Page 3

```
APPLICANT: ORSINI, Cecile  
; APPLICANT: YEH, Patrice  
APPLICANT: LATTA, Martine  
APPLICANT: PROST, Edouard  
TITLE OF INVENTION: CELLS FOR THE PRODUCTION OF RECOMBINANT  
NUMBER OF SEQUENCES: 8  
CORRESPONDENCE ADDRESS:  
ADDRESSER: Rhone-Poulenc Rorer Inc.  
STREET: 500 Arcoia Road, Mailstop 3c43  
CITY: Collegeville  
STATE: PA  
COUNTRY: USA  
ZIP: 19426
```

```
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/M5-  
SOFTWARE: Patentin Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/875,223  
  
FILING DATE:
```

```
CLASSIFICATION:  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 95/00747  
FILING DATE: 20-JAN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 95/06532  
FILING DATE: 01-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR 95/10541  
FILING DATE: 08-SEP-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: FR WO F96/00088  
FILING DATE: 19-JAN-1996  
ATTORNEY/AGENT INFORMATION:  
NAME: Fehlner Esq., Paul F.  
REGISTRATION NUMBER: 35,135  
REFERENCE/DOCKET NUMBER: ST9500561-US  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (610) 454-3839  
TELEFAX: (610) 454-3808  
INFORMATION FOR SEQ ID NO.: 8:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 534 amino acids  
TYPE: amino acid  
STRANDNESS: single  
TOPOLOGY: linear  
MOLECULE TYPE: protein  
US-08-875-223-8
```

```
Query Match      6.5%; Score 68.5; DB 4; Length 534;  
Best Local Similarity 23.9%; Pred. No. 10;  
Matches    26; Conservative   21; Mismatches     27; Indels       35; Gaps        6.
```

```
OY          25 INTMTSLNNGRYAGGYCMASDRIGDSYC-AGRMGVTLFSRSBYAIFEVKLKNAAYLI 83  
           |:|::||::||:::||::|:|:|:|:|:|:|:|:|:|:|:~::~||:  
Db         346 IIDTSMTHIGHESMAHCHCSS-FGSIQCTIAGGGVLASF-RMV-----DGAMF-- 394
```

```
OY          84 WLGIQRAGAIDLKSIASTSRHLFORAVFNLTPRKSVITLALF 132  
           ::::| | || ~~~~~~NNRIWRREV-VANNPKREVMSSVF 420  
Db         395 -----
```

```
RESULT      6  
US-08-448-603a-30  
Sequence 30, Application US/08448603a  
Patent No. 5864027  
GENERAL INFORMATION:  
APPLICANT: Berkman, Phillip W.  
APPLICANT: Nakamura, Gerald R.  
TITLE OF INVENTION: HIV Envelope Polypeptides  
;
```

```

1 NUMBER OF SEQUENCES: 33
2 CORRESPONDENCE ADDRESS:
3 ADDRESSEE: McCutchen, Doyle, Brown & Eneisen, LLP
4 STREET: 3 Embarcadero Center
5 City: San Francisco
6 STATE: CA
7 COUNTRY: USA
8
9 ZIP: 94111
10
11 COMPUTER READABLE FORM:
12 MEDIUM TYPE: Diskette
13 COMPUTER: IBM Compatible
14 OPERATING SYSTEM: DOS
15 SOFTWARE: FASTSEQ for Windows Version 2.0
16
17 CURRENT APPLICATION DATA:
18 APPLICATION NUMBER: US/08/448,603A
19 FILING DATE: 07-JUN-1994
20 CLASSIFICATION: 435
21
22 PRIOR APPLICATION DATA:
23 APPLICATION NUMBER: 08/072,833
24 FILING DATE: 07-JUN-93
25 ATTORNEY/AGENT INFORMATION:
26 NAME: Haliday, Emily
27 REGISTRATION NUMBER: 38903
28 REFERENCE/DOCKET NUMBER: 14918-704
29 TELECOMMUNICATION INFORMATION:
30 TELEPHONE: 415-393-2000
31 TELEFAX: 415-393-2286
32
33 TELEX:
34
35 INFORMATION FOR SEQ ID NO: 30:
36 SEQUENCE CHARACTERISTICS:
37 LENGTH: 857 amino acids
38 TYPE: amino acid
39 STRANDEDNESS: single
40 TOPOLOGY: linear
41
42 US-08-448-603A-30
43
44
45
46
47
48
49
50
51
52
53
54
55
56
57
58
59
60
61
62
63
64
65
66
67
68
69
70
71
72
73
74
75
76
77
78
79
80
81
82
83
84
85
86
87
88
89
90
91
92
93
94
95
96
97
98
99
100

```


Db 844 IFPVCSITGILLAHVAMKIAVLL-AAVPVMITAGYRLRLVLAESRRHSAYNDAA 902
 Qy 116 EVNLTKNSIVFLALFPQFIMPOQPOLMOYIVLGT-----TIYVDI 158
 Db 903 SIAEACRGIRTASIGRE-----RGVSRASNAVPEPYDKGIRFLTLINT 948
 Qy 159 IVMIGYAT--LAORIALMWIKPKOMKALNKIFGSL--FMVGCALLASARHA 205
 Db 949 LIALSEITFYVALAW WGAQQR--NGTYSQLEDFIVLALLFSQAOSA 996

RESULT 10
 US-08-358-171-2
 ; Sequence 2, Application US/08358171
 ; Patent No. 5763578
 ; GENERAL INFORMATION:
 ; APPLICANT: FONG, Henry K.W.
 ; TITLE OF INVENTION: ALK TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
 ; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/358,171
 ; FILING DATE: 16-DEC-1994
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: YUN, Allen C.
 ; REGISTRATION NUMBER: 37,971
 ; REFERENCE/DOCKET NUMBER: FONG-2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528
 ; TELEX: 248633
 ; INFORMATION FOR SEQ. ID NO.: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 291 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-08-358-171-2

Query Match 6.3%; Score 66.5; DB 1; Length 291;
 Best Local Similarity 21.7%; Pred. No. 7.4;
 Matches 51; Conservative 35; Mismatches 76; Indels 73; Gaps 13;
 Qy 11 LTSILTLSPGSGAINM--TSLNHGYPAGGYCWMASDRGDSYCAWMRG---VGTU 63
 Db 53 LTVSLALADSGISLNLVAATSSLLRMPYSGEQAH-----GFGFVALASI 103
 Qy 64 FRSRVIAF-----EVLKWAQA--AYLIWLGIOQWRAA-----GAILDKSLAS--- 103
 Db 104 CSSAAVAMGRYHHCCTRSRLDMNTAVSLVEFWLSAFAWALPLLGWGHYDEPLGTCT 163
 Qy 104 ---TQSRRLHFGRAVFNVLNPKSIYFLAALFPQFIMPOQPOLMOYIVG-----VTI 154
 Db 164 LDYSRGDRN-FTSFLF-----TMAFENFLDLFTTVVSYRLMDQ-KLGKTSRPVNTV 214
 Qy 155 VVDIIVMIG-----YATLAQRI-----ALMIKPKOMKALNKIFGS 190
 Db 215 LPARTLLIGWGPYALLVYVATITADATISPKLQMWPALIAKAVPTVMANVVALGS 269

RESULT 11
 US-09-090-947-2
 ; Sequence 2, Application US/09090947
 ; Patent No. 6008338
 ; GENERAL INFORMATION:
 ; APPLICANT: FONG, Henry K.W.
 ; TITLE OF INVENTION: ALK TRANS-RETINALDEHYDE BINDING PROTEIN, DNA
 ; TITLE OF INVENTION: ENCODING SAME, AND ANTIBODIES THERETO
 ; NUMBER OF SEQUENCES: 25
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: BROWDY AND NEIMARK
 ; STREET: 419 Seventh Street, N.W., Suite 300
 ; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: USA
 ; ZIP: 20004
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: Patentin Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/090,947
 ; FILING DATE:
 ; CLASSIFICATION:
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/358,171
 ; FILING DATE:
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: YUN, Allen C.
 ; REGISTRATION NUMBER: 37,971
 ; REFERENCE/DOCKET NUMBER: FONG-2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-628-5197
 ; TELEFAX: 202-737-3528
 ; TELEX: 248633
 ; INFORMATION FOR SEQ. ID NO.: 2:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 291 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; US-09-090-947-2

Query Match 6.3%; Score 66.5; DB 3; Length 291;
 Best Local Similarity 21.7%; Pred. No. 7.4;
 Matches 51; Conservative 35; Mismatches 76; Indels 73; Gaps 13;
 Qy 11 LTSILTLSPGSGAINM--TSLNHGYPAGGYCWMASDRGDSYCAWMRG---VGTU 63
 Db 53 LTVSLALADSGISLNLVAATSSLLRMPYSGEQAH-----GFGFVALASI 103
 Qy 64 FRSRVIAF-----EVLKWAQA--AYLIWLGIOQWRAA-----GAILDKSLAS--- 103
 Db 104 CSSAAVAMGRYHHCCTRSRLDMNTAVSLVEFWLSAFAWALPLLGWGHYDEPLGTCT 163
 Qy 104 ---TQSRRLHFGRAVFNVLNPKSIYFLAALFPQFIMPOQPOLMOYIVG-----VTI 154
 Db 164 LDYSRGDRN-FTSFLF-----TMAFENFLDLFTTVVSYRLMDQ-KLGKTSRPVNTV 214
 Qy 155 VVDIIVMIG-----YATLAQRI-----ALMIKPKOMKALNKIFGS 190
 Db 215 LPARTLLIGWGPYALLVYVATITADATISPKLQMWPALIAKAVPTVMANVVALGS 269

RESULT 12
 US-08-468-812-2
 ; Sequence 2, Application US/08468812
 ; Patent No. 5935836
 ; GENERAL INFORMATION:
 ; APPLICANT: Vehmaanper, Jari

```

; APPLICANT: M ntyl, Arja
; APPLICANT: Fagerstr m, Richard
; APPLICANT: Iantto, Raija
; APPLICANT: Palohelmo, Marja
; APPLICANT: Suominen, Pirkko
; APPLICANT: Lahtinen, Tarja
; APPLICANT: Kristo, Paula
; TITLE OF INVENTION: Actinomadura Xylanase Sequences and Methods
; TITLE OF INVENTION: of use
; NUMBER OF SEQUENCES: 25
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: STERN, KESSLER, GOLSTEIN & FOX
; STREET: 1100 New York Ave., N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.A.
; ZIP: 20005
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/468,812
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/332,412
; FILING DATE: 31-OCT-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/282,001
; FILING DATE: 29-JUL-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Bugalsky, Larry B.
; REGISTRATION NUMBER: 35,086
; REFERENCE/DOCKET NUMBER: 1050.0340002
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-371-2600
; TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 344 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-468-812-2

Query Match 6.3%; Score 66.5; DB 2; Length 344;
Best Local Similarity 22.7%; Pred. No. 9.3;
Matches 25; Conservative 11; Mismatches 45; Indels 29; Gaps 3;

QY 10 LLSITLSPGSGAINTMTSLNHGYAGVYCWASD-----RTGSDYCGMRGV 60
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 28 LALAIGALLPGTAHDTTITONOTGNDNGYRYSWTDAFGVSKTLHSGGSYSTSWRNT 87
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 GTLFS-----RSVIAFEVLKWAAYLIWLG-----IQQW 90
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 88 GNFVAGKMGSTGGRRTVYNASFNPSGNYLTLXGWTNPLVEYYIYVSW 137
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 13
US-08-863-102-1
; Sequence 1, Application US/08863102
; Patent No. 6015828
; GENERAL INFORMATION:
; APPLICANT: Cuppoletti, John
; TITLE OF INVENTION: Chemical Modification of
; TITLE OF INVENTION: Chloride Channels as a Treatment for Cystic
; TITLE OF INVENTION: Fibrosis and Other Diseases
; NUMBER OF SEQUENCES: 4
```

```

; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jackie A. Zurcher,
; ADDRESSEE: Dinmore & Shohl LLP
; STREET: 255 E. Fifth Street,
; STREET: 1900 Chemed Center
; CITY: Cincinnati
; STATE: Ohio
; COUNTRY: USA
; ZIP: 45202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.5 inch
; COMPUTER: IBM PC Compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/863,102
; FILING DATE: 23-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Zurcher, J. A.
; REGISTRATION NUMBER: 42,251
; REFERENCE/DOCKET NUMBER: 10738-12
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (513) 977-8377
; TELEFAX: (513) 977-8141
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 898 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; FEATURE:
; NAME/KEY: Rabbit Gastric CIC-2
; US-08-863-102-1

Query Match 6.3%; Score 66.5; DB 3; Length 898;
Best Local Similarity 25.0%; Pred. No. 35;
Matches 47; Conservative 26; Mismatches 76; Indels 39; Gaps 10;

QY 39 GGVYCW---ASDRT-----GDSYCGMRGVGTLFRSVIAFEVLKWAAYLIWLGIO 88
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 47 GGEPEWRSPSPRPPELLEYGQSRCAKRCMSYRCHKFLVSRGDEW---ITLVLGIL 103
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 89 QWRAGAIDLKSLASTOSRHLFORAVFVNL-----TNPSIYFLAALFPQFIIMPQ 140
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 104 MALVSMADYALAACLOQOWM-SRGINTNLLQYLAWTVYPPVLIFFSAGFTQILAPQ 162
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 141 -----PQIMQYIVIGVT-----TIVVDIYVIGY-ATLAQRILMIKGP-----KQMKAL 184
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 163 VQSGTPE-MKTLRGVVLKEYTLTKTFVAKYIGLTCALGSGMPLGKGPVHTASMCAL 221
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 185 NKIFGSLF 192
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 222 LSKFLSLF 229
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |

RESULT 14
US-08-863-102-4
; Sequence 4, Application US/08863102
; Patent No. 6015828
; GENERAL INFORMATION:
; APPLICANT: Cuppoletti, John
; TITLE OF INVENTION: Chemical Modification of
; TITLE OF INVENTION: Chloride Channels as a Treatment for Cystic
; TITLE OF INVENTION: Fibrosis and Other Diseases
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Jackie A. Zurcher,
; ADDRESSEE: Dinmore & Shohl LLP
; STREET: 255 E. Fifth Street,
; STREET: 1900 Chemed Center
; CITY: Cincinnati
; STATE: Ohio
```

```

: COUNTRY: USA
: ZIP: 45202
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette, 3.5 inch
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: MS-DOS
: SOFTWARE: WordPerfect 6.1
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/863,102
: FILING DATE: 23-MAY-1997
: ATTORNEY/AGENT INFORMATION:
: NAME: Zurcher, J. A.
: REGISTRATION NUMBER: 42,251
: REFERENCE/DOCKET NUMBER: 10738-12
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (513) 977-8377
: TELEFAX: (513) 977-8141
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 898 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: FEATURE:
: NAME/KEY: Human Gastric CIC-2
: US-08-863-102-4

```

```

Query Match 6.3%; Score 66.5; DB 3; Length 898;
Best Local Similarity 25.0%; Pred. No. 35;
Matches 47; Conservative 26; Mismatches 76; Indels 39; Gaps 10;

QY 39 GGVYQW---ASDRT-----GDSYCACMRGVGTLFSRSVIAFEVLMKAGAVLIMLQ 88
   || || || || || || || || || || || || || || || || || || || ||
Db 47 GGPWRSPSPRTPELLEFGSCARCRCMSVCHKEFVSRRGEDM---IFVLGLL 103
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 89 QWRAGATIDKSLASTOSRRHLEFORAVFVN-----TNPKSVFLAALFPQFIMQO 140
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 104 MALVSWAMDYAIACILQAOOWM-SRGLNTNLLQYLAVVYPPVVLITPSAGFTQILAPQ 162
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 141 -----PQIMQYIVLGV-----TIIVDLIMVGY-ATLAORIALWIKGP-----KOMAL 184
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 163 VSGGIPK-MKTLIRGVLYKEVLTFTVAKVIGLICALGSGMPKREPFVHIAMCAL 221
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 185 NKIFGSLF 192
   ||||
Db 222 LSKFLSLF 229

```

```

RESULT 15
US-08-802-141-4
: Sequence 4, Application US/08802141
: Patent No. 5773009
: GENERAL INFORMATION:
: APPLICANT: GLASS, ROGER I.
: APPLICANT: GENTSCH, JOHN R.
: APPLICANT: BHAN, M. K.
: APPLICANT: DAS, BIMAL K.
: TITLE OF INVENTION: ROTAVIRUS STRAIN AND RELATED
: TITLE OF INVENTION: COMPOSITIONS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: NEEDLE & ROSENBERG, P.C.
: STREET: 127 Peachtree Street, Suite 1200
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25

```

```

: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/802,141
: FILING DATE: 19-FEB-1997
: CLASSIFICATION: 424
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 08/231,041
: FILING DATE: 15-APR-1994
: ATTORNEY/AGENT INFORMATION:
: NAME: Spratt, Gwendolyn D.
: REGISTRATION NUMBER: 36,016
: REFERENCE/DOCKET NUMBER: 1414.609
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 404/688-0770
: TELEFAX: 688-9880
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 772 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-802-141-4

```

```

Query Match 6.2%; Score 66; DB 1; Length 772;
Best Local Similarity 20.0%; Pred. No. 33;
Matches 38; Conservative 28; Mismatches 58; Indels 66; Gaps 8;

QY 5 WMPAYLLTSLITLSPGSGAINTMTTSLNHGYPAGVYCMASDRTGDSYCACMRGVGTLF 64
   || || || || || || || || || || || || || || || || || || || ||
Db 102 WMFTYL---VLPMTAQTSVVYVNMNETVN-----ISIDNSGAV----- 137
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 65 SRVIAFEVLMKAGAVLIMLGIQWRAGATIDKSLASTOSRRHLEFORAVFVNLTN-- 121
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 138 -KEVDYFKT-----SSAOAYSRNPLITAHRLQAYKKDGGGNSNYNG 179
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 122 -----PKSIVFLAALFPQFIMP--QQPQIMQYIVLGVTTI-VVDIIVAI 162
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
Db 180 SDAYGDLRVGTYFNPNVNAVINLADF--YVIPDSQOEMCTEYIRGLPAIQTTTYVTP 237
   :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :| :|
QY 163 GYATLAQRIA 172
   ||||
Db 238 SYAVRSORIA 247

```

Search completed: May 6, 2001, 14:39:23
Job time: 414 sec

OM of: US-09-466-935-2 to: GenEmbl:* out_format : pfs
Date: May 6, 2001 3:38 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODE=frame+pn.model -DEV=xlp
-Q/cg2.1/USPTO_spool/US09466935/runat_03052001.075709_14798/app_query.fasta_1.529
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOCPCL=0.000 -LOOEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
-FCAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELCP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALLIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09466935.@CGN1.1 5396 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-466-935-2
Query length: 205
Database: GenEmbl:*
Database sequences: 1283235
Database length: -1216004940
Search time (sec): 1745.950000

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pat1:AX030175	+	1061.00	1875.45	3.0e-96	1231	AX030175 Sequence 1 from Paten
gb_pat2:ECOM85	+	1023.00	1796.54	7.3e-92	91414	M87049 E. coli genomic sequen
gb_pat1:AE000458	-	1024.00	1788.94	2.0e-91	11509	AE000458 Escherichia coli K12
gb_pat1:AE0005613	-	950.00	1657.93	3.9e-84	10592	AE0005613 Escherichia coli O15
gb_pat3:STSTMTM1	-	845.00	1450.61	1.4e-72	96086	AF233324 Salmonella typhimur
gb_htg9:AC020870	-	795.00	1352.52	4.0e-67	256373	AC020870 Mus musculus clone
gb_htg9:AC022157	-	751.50	1311.69	7.6e-65	4883	AC022157 Homo sapiens chromosc
gb_htg9:AC020970	-	734.00	1245.60	3.7e-61	212936	AC020970 Mus musculus clone
gb_htg7:AC020874	+	637.00	1070.84	2.0e-51	265383	AC020874 Mus musculus clone
gb_htg11:AC026742	-	477.00	790.02	8.7e-36	170178	AC026742 Homo sapiens chrom
gb_pat1:AE004104	-	460.50	789.21	9.6e-36	11305	AE004104 Vibrio cholerae chrc
gb_pat1:AE0004937	-	402.50	679.51	1.2e-29	10475	AE0004937 Pseudomonas aerugin
gb_htg7:AC020885	+	397.50	644.25	1.1e-27	268294	AC020885 Mus musculus clone
gb_htg9:AC020833	+	270.50	431.56	8.0e-16	62274	AC020833 Mus musculus clone
gb_pat2:AP003001	+	255.50	389.04	1.9e-13	345783	AP003001 Mesorhizobium loti
gb_pat2:AP003005	+	235.50	353.36	1.8e-11	347950	AP003013 Mesorhizobium loti
gb_pat2:BSDB0014	-	228.50	341.31	8.5e-11	332635	AP003005 Mesorhizobium loti
gb_pat1:AE0002037	-	228.00	370.87	1.9e-12	12268	AE0002037 Deinococcus radiodur
gb_pat2:AP000295	-	218.50	323.12	8.8e-10	346897	AP000295 Mesorhizobium loti
gb_pat1:AF157493	+	203.50	320.51	1.2e-09	25454	AF157493 Zymomonas mobilis ZM
gb_pat2:AP003012	-	200.50	291.07	5.4e-08	346547	AP003012 Mesorhizobium loti
gb_pat2:BSDB0014	-	197.00	289.31	6.7e-08	213420	Z59117 Bacillus subtilis com
gb_pat1:AX030081	+	195.50	340.25	9.8e-11	639	AX030081 Sequence 11 from Paten
gb_pat1:AE000274	-	195.50	311.91	3.7e-09	13793	AE000274 Escherichia coli K12
gb_pat2:D90824	-	195.50	311.62	3.9e-09	14236	D90824 E. coli genomic DNA, KC
gb_pat2:ID90823	-	195.50	310.85	4.3e-09	15475	D90823 E. coli genomic DNA, KC
gb_pat1:AB016260	-	195.50	286.95	9.1e-08	204479	AB016260 Agrobacterium tumef
gb_pat1:AE0004864	-	193.00	309.57	5.0e-09	10976	AE0004864 Pseudomonas aerugin
gb_pat2:AP003008	+	193.00	277.71	3.0e-07	346820	AP003008 Mesorhizobium loti
gb_pat2:AP001568	+	190.50	274.66	4.4e-07	296950	AP001508 Bacillus halodurans
gb_pat1:AE0004221	+	190.00	303.87	1.0e-08	11401	AE0004221 Vibrio cholerae chrc
gb_pat2:AP003004	-	189.50	271.70	6.4e-07	338579	AP003004 Mesorhizobium loti
gb_pat2:AP003012	+	187.50	267.93	1.0e-06	346547	AP003012 Mesorhizobium loti
gb_pat1:AE0004372	+	186.50	298.64	2.0e-08	10176	AE0004372 Vibrio cholerae chrc
gb_pat1:AE0004786	-	186.50	298.64	2.0e-08	10235	AE0004786 Pseudomonas aerugin
gb_pat1:AE0005402	-	186.50	297.71	2.3e-08	11313	AE0005402 Escherichia coli O15
gb_pat1:AB024601	-	186.50	295.23	3.2e-08	14807	AB024601 Pseudomonas aerugin
gb_htg25:LMFICHR32.20	+	186.00	275.84	3.8e-07	110000	Continuation (21 of 28) of
gb_pat2:AR061372	+	184.00	316.77	2.0e-09	720	AE061372 Sequence 6 from patent
gb_pat3:SCMIGA	+	184.00	316.77	2.0e-09	885	X67020 S.colwelliana mlgA gene
gb_pat1:AF188935	+	184.00	273.51	5.1e-07	96231	AF188935 Bacillus anthracis p

gb_pat1:AE004889	-	180.50	286.95	9.1e-08	11415	AE004889 Pseudomonas aerug
gb_pat1:AE004079	+	179.00	285.09	1.2e-07	10445	AE004079 xyliella fastidios
gb_pat3:PA081259	+	178.50	297.46	2.4e-08	2483	U04992 Pseudomonas aerugin
gb_pat3:PA081259	+	178.50	287.53	8.5e-08	7285	U81259 Pseudomonas aerugin

seq_name: gb_pat1:AX030175

seq_documentation_block:

LOCUS AX030175 1231 bp DNA
DEFINITION Sequence 1 from Patent EP1013765.
ACCESSION AX030175
VERSION AX030175.1 GI:10190392
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
1 (bases 1 to 1231)
Belareva,A.V., Aleshin,V.V., Livshits,V.A., Tokmakova,I.L. and

TITLE
Gene and method for producing l-amino acids
JOURNAL
Patent: EP 1013765-A 1 28-JUN-2000;
AJINOMOTO KK (JP)

FEATURES
location/Qualifiers

source
1..1231
/organism="Escherichia coli"
/db_xref="taxon:562"
557..1174
/note="unnamed protein product"
/codon_start=1
/transl_table=11
/protein_id="CAC09318.1"
/db_xref="GI:10190393"
/translation="MTLEWFFAVLLTSLTSPGSAINTMTTSLNHGYPAGGYCM
ASDPTGDSYAGWGVGTLSRSYIAEVEIKMGAAFLIMGLDQWMAAGADIDLSIA
STOSRRLFGAVVNLNPKRSIVFLALPPFPOOLMOYIVIGVTVIVDITV
MIGVATLAORIALIMIKGPKMKMLNKFESLFLMVGALLASARHA"

BASE COUNT
291 a 319 c 311 g 310 t
ORIGIN

alignment_scores:
Quality: 1061.00 Length: 205
Ratio: 5.176 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000

alignment_block:
US-09-466-935-2 x AX030175 ..
Align seg 1/1 to: AX030175 from: 1 to: 1231

1	MetThrLeuGluTrpTrpPheAlaTyrLeuLeuThrSerIleLeu	17
557	ATGACCTTGAATAGTGGTTGGCTTGCATCGACATCATTTTAA	606
17	riLeuSerProGlySerClyAlaIleAsnThrMetThrThrSerLeu	34
607	GGCTGTGCCAGGCTGTGGCAATCAACATATGACACCTGCTCAAC	656
34	isGlyTyrProAlaGlyGlyValTyrCysTrpAlaSerAspArgTrg	50
657	ACGGTATTCGGCCGGTGGCTGCTATTCGTGGCTTCAACCGGACTGC	706
51	AspSerTyrCysAlaGlyTyrTrpArgGlyValGlyThrLeuPheSer	67
707	GATTCATATTTGGTGGTGGCTGGGCTTGGAGCGATTTTCCGGTC	756
67	TrAlIleAlaPheGlyValLeuLeuTyrTrpAlaGlyAlaTyrLeu	84
757	AGTGATTTGGCTTGAAGTCTGAAGTGGCGGCGGCTTACTGATT	806
84	TrLeuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspLeuYss	100

807 GGCTGGAAATCCAGCACTGGCGCGCGCTGGTCAATTGACCTTAATCG 856
101 LeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValPheVa 117
|||||
857 GTGGCCCTCAATCAATCGCGTCGACATTTGTTCCAGCGCCAGTTTGT 906
117 IAsnLeuThrAsnProLysSerIleValPheLeuAlaIleLeuPheProG 134
|||||
907 GAATCTACCAATCCCAAAAGATATTGTTTCGGCGCGCTATTTCGCG 956
134 InpIleIleMetProGlnGlnProGlnLeuMetGlnTrpIleValLeuGly 150
|||||
957 AATTCATCATGATCGCGACAGCCGCACTGATGCATGATATCGCTCGGC 1006
151 ValThrThrIleValAlaAspIleIleValMetIleGlyTrpAlaThrIle 167
|||||
1007 GTCCACCACTATTGGTCGATATTATTGTGATGATCGCTTACGCCACCT 1056
167 uAlaGlnArgIleAlaLeuTrpIleLysGlyProLysGlnMetLysAlaI 184
|||||
1057 TGCACAACGAGATTGCTCTATGATTAAGACCAAGCAGATGAAGCGCG 1106
184 eAsnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeuLeuAla 200
|||||
1107 TGATATAGATTTCGCTGCTGTTATGCTGGTGAGCGCTGTAGCA 1156
201 SerAlaArgHisAla 205
|||||
1157 TCGCGAGGCGCATGCG 1171

seq_name: gb_ba2:ECOW85

seq_documentation block:

LOCUS ECOW85 91414 bp DNA BCT 29-MAY-1995
DEFINITION E. coli genomic sequence of the region from 84.5 to 86.5 minutes.
ACCESSION M87049
VERSION M87049.1 GI:836656
KEYWORDS
SOURCE
ORGANISM
Escherichia coli.
Escherichia coli.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE
AUTHORS 1 (bases 1 to 91414)
TITLE Daniels,D.L., Plunkett,G. III., Burland,V. and Blattner,F.R.
JOURNAL Analysis of the Escherichia coli genome: DNA sequence of the region
MEDLINE Science 257 (5071), 771-778 (1992)
REFERENCE
AUTHORS 2 (bases 1 to 91414)
TITLE Plunkett,G. III., Burland,V., Daniels,D.L. and Blattner,F.R.
JOURNAL Analysis of the Escherichia coli genome. III. DNA sequence of the
MEDLINE region from 87.2 to 89.2 minutes
REFERENCE
AUTHORS 3 (bases 1 to 91414)
TITLE Blattner,F.R., Burland,V., Plunkett,G. III., Sofia,H.J. and
JOURNAL Daniels,D.L.
MEDLINE Analysis of the Escherichia coli genome. IV. DNA sequence of the
REFERENCE region from 89.2 to 92.8 minutes
AUTHORS 4 (bases 1 to 91414)
TITLE Rudde,K.E., Sofia,H.J., Koonin,E.V., Plunkett,G. 3rd, Lazar,S. and
JOURNAL Rouviere,P.E.
MEDLINE A new family of peptidyl-prolyl isomerases
REFERENCE 5 (bases 1 to 91414)
AUTHORS Daniels,D.L.
JOURNAL Trends Biochem. Sci. 20 (1), 12-14 (1995)
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1992) Donna L. Daniels, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecollegenetics.wisc.edu; Phone: 608-262-2534; Fax:

COMMENT

608-263-7459
On May 29, 1995 this sequence version replaced gi:148169.
This sequence was determined as part of the E. coli Genome Project
(Frederick R. Blattner, director) at the university of
Wisconsin-Madison. The entire sequence was independently determined
from E. coli MG1665. Overlaps and conflicts with other sequence
determinations are annotated. Reference [1] describes the original
sequence determination of a 91408 bp sequence. References [2], [3],
and [4] describe subsequent corrections and/or updates to that
sequence.

FEATURES

source

1..91414
/organism="Escherichia coli"
/strain="K-12"
/db_xref="taxon:562"
/clone_lib="lambda library of Daniels and Blattner"
/map_bp 1 at 3974 kb: 84.7 minutes"
/note="This sequence comprises the following lambda
clones: DD850(EC15-116), DD851(EC13RM4), DD854(EC15-177),
DD856(EC19RM8.1), DD858(EC17RM9),
DD861(EC17-221), DD864(EC27-149), DD865(EC21-76),
DD867(EC27-860), M13mp19 or Janus vectors were used for
subcloning."
1..200
/note="corresponds to K02846 (1..200)"
1..5863
/note="clone EC15-116"
complement(1..4899)
/note="corresponds to M37337 (1..4900)"
1..339
/note="corresponds to V00349 (164..500)"
10..86
/gene="aspr"
/note="CGSC No. 989"
10..86
/gene="aspr"
95..170
/gene="trpP"
/note="CGSC No. 66"
95..170
/gene="trpP"
177..205
/note="terminator-like sequence"
253..396
/note="predicted bend of 81 degrees"
complement(266..862)
/gene="f198"
complement(266..862)
/gene="f198"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AA067566.1"
/db_xref="GI:148170"
/translation="MAHTSRHNEFSIGASASLMECLNQMIGRLYONODAHFLOFEA
RLAOROSLVKOLHEAGLIGADVEPIITTSASAEIAOQIAMLNGCTWLPVSMARKKGLHT
VDSFTLSRPYALWLNQSDKNALRDILKINLDEY"
586..724
/note="predicted bend of 77 degrees"
complement(703..1104)
/gene="f133"
complement(703..1104)
/gene="f133"
/note="CGSC No. 18010; possible psr gene"
/codon_start=1
/transl_table=11
/protein_id="AA067567.1"
/db_xref="GI:148171"
/translation="MDTELKTFLEVSTRHFGRAESLYLQSAVSRIRLOENQIG
VNLFTFRNNRIRLTAAGERLALPYAETLWSTQWAAARKRWIRIPDITISLVPAKRGNV
CLISGWDACIRIRMPIDQAYSSKRELNGSLW"

promoter complement(1136. .1165)
/note="promoter matrix score of 63"
1151. .1179
promoter /note="promoter matrix score of 57"
1223. .1636
gene /gene="0137"
1223. .1636
CDS /gene="0137"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AA067568.1"
/db_xref="GI:148172"
/translation="MAESFTTNRYFNKHYPRGFSRKGDTIKEAOLLEHCGAENE
LDGKREPVTEEEKLEVAVCGEREPEVTEARVSKITWTRIKRKRPHHTLSAVNRRLK
VLKTPILITINKKMPFVYSSFCRMMSKRSIAR"
1562. .1586
terminator /gene="0137"
/note="terminator-like sequence"
complement(1585. .3135)
gene /gene="F516"
complement(1585. .3135)
CDS /gene="F516"
/note="similar to Mg chelatase subunit of Rhodobacter
capsulata"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AA067569.1"
/db_xref="GI:148173"
/translation="MHIMVTOGGLMSTLSIVHTRAALGNAPRTVEVHISGLPLIM
VGLPRTYKARDRYRSALINSGYEPRAKTITINLAPRDLKEGGRDLPFLATILLA
SGLTANKUDEVELYELALTGALRGVPAKTSATKATSGRKTITVADNDEGLIN
SGGCIADHDLAVCAFLGKHALEKPKPDVAISRALQHLSDVIGQEGKGLDITTA
GGHNLILIGPQGTMTLASRINGLIPDLSDEALIESAAILSVNAESVQKQMPORF
RSPHNSASLTAAVGGGAIIPGGEISLANGVLEFDELPEFRRTLDALREPTEGGQIL
LSRTAKITTPARFQTLVAAMPSPITGHYOGHNHRCPTQTLRYLNRISGPLDRFDS
LEIPLPQGLISKTYVPGSSATVQRYMARERFRKONLNMWDSPEFRQCKLE
SEDAWMBSTLHLGLSTRAWORLTKVARTIADDSOITIRHQLQEAVSYRAIDRL
IHLQKILT"
misc_structure 3287. .3433
/note="predicted bend of 79 degrees"
3297. .3326
promoter /note="promoter matrix score of 74"
3336. .3676
misc_feature /note="corresponds to M11651 (1. .341)"
3336. .3737
misc_feature /note="corresponds to V00289 (1. .402)"
3336. .12791
misc_feature /note="corresponds to M10313 (1. .9456)"
3379. .5721
gene /note="V00290; ECILVX(1. .2343)"
3458. .3556
gene /gene="11VL"
3458. .3556
CDS /gene="11VL"
/codon_start=1
/transl_table=11
/product="l1vGMDA operon leader peptide"
/protein_id="AA067570.1"
/db_xref="GI:148174"
/translation="MTALLRVISLVVISVYIIIPCCGALGRGKA"
3696. .4679
gene /gene="11vG"
3696. .4679
CDS /gene="11vG"
/note="CGSC No. 603; inactive in wild-type E. coli K12"
/codon_start=1
/transl_table=11
/product="acetohydroxy acid synthase II, large subunit"
/protein_id="AA067571.1"
/db_xref="GI:148175"
/translation="MNGAQMVVYHALRAGCVNTVFGYPGGAIIMPYDALYDGGVEHLIC

RHOGAAMAIGARATGKTGVCIAATSGPATNLITGLADALIDSIPVATIGQVSAP
FICTDARFOEDVINDICTAKHSFIVOSLEELPRIMAEAFVAGSGRGPVVDIPKD
IQLASDLEPWTFTTVENEYXPHAEVEQAROKAROKMLYGGGVGAQAVPILRE
FLAATMPATCTYGLGAVAEADPYTLGMLGMHGTAAANFAVQBCDLIAVGARDDR
VTKGLNTPAPHASVITHMDIDPAEMNKLROAHVALQGLNALPALQOPLNO"
gene 4675. .5340
/gene="0221"
4675. .5340
CDS /gene="0221"
/note="the mutation l1v02096 joins this ORF to l1vG,
restoring function missing in wild-type"
/pseudo
/codon_start=1
variation replace(4677. .4678, "tatg")
alignment_scores:
Quality: 1039.00 Length: 206
Ratio: 5.093 Gaps: 1
Percent Similarity: 99.029 Percent Identity: 99.029
alignment_block:
US-09-466-935-2 x ECOUW85/rev ..
Align seg 1/1 to reverse of: ECOUW85 from: 1 to: 91414
1 MetThrLeuGluTrpTrpPheLeuAlaTrpLeuThrSerIleLeuLeu 17
|||||
62161 ATGACCTTAGAATGGTGGTGGCTTACCTGCGACATGATCATTTTATAC 62112
17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrSerLeuAsnH 34
|||||
62111 GGTGTGCCAGGCTCTGGTGCATCAACACTATAGCACCACTGCGTCAACC 62062
34 IScGlyTyr. ProAlaGlyGlyValIlyrcystrpAlaSeraspArgThrG1 50
|||||
62061 ACGGTATACSCGCGGCGGCGCTCTATGCTGCGCTTACGACCGGACCTGG 62012
50 yAspSerTyrCysAlaGlyTrpArgGlyValGlyThrLeuPheSerArgS 67
|||||
62011 CCAATTCATATTTGCTGGTGGTGGGCGGCGGCGCATTTTTCGGCT 61962
67 erValIleAlaIlePheGluValLeuLysTrpAlaGlyAlaAlaTrpLeuIle 83
|||||
61961 CAGTATTCGCTTGAAGCTTGAAGTGGCGAGCGCGGCTTACTTGATT 61912
84 TrpLeuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspLeuLysSe 100
|||||
61911 TGGCTGGGAATCCAGACAGTGGCGCGCGCTGCGCAATTGACCTTAAATC 61862
100 rLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValPheY 117
|||||
61861 GGTGGCTTACTCAATCGCTGCACATTGTTCACAGCGCGGAGTTTGTG 61812
117 aLAsnLeuThrAsnProLysSerIleValPheLeuAlaIleLeuPhePro 133
|||||
61811 TGAATTCACCAATCCCAAAAGTATTTGTTCTGGCGGCGCTATTTCCG 61762
134 GlnPheIleMetProGlnGlnProGlnLeuMetGlnTyrIleValLeuG1 150
|||||
61761 CAATTCATCATGCGCGACACGCGCAACTGATGATATATGCTGCTGG 61712
150 yValThrThrIleValIleAspIleIleValMetIleGlyTyrAlaThrL 167
|||||
61711 CGTCACACTATTTGGTGGATATTTGTGATGATGGTGTACGCGCACCC 61662
167 euAlaGlnArgIleAlaLeuTrpIleLysGlyProLysGlnMetLysAla 183
|||||
61661 TTGCTCAACGATGCTCTATGATTTAAAGACCAAGACAGTGAAGCGG 61612
184 LeuAsnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeuLeuAl 200
|||||
61611 CTGAATTAAGATTTTGCGGCTGTTGTTATGCTGGTGGAGCGGCTGTATAC 61562

200 aSeraLaarGHisAla 205
|||||
61561 ATCGCGCAGCATGCG 61546

seq_name: gb_ba1:AE000458

seq_documentation_block:
LOCUS AE000458 11509 bp DNA BCT 01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 348 of 400 of the complete genome
ACCESSION AE000458 000096
VERSION AE000458.1 GI:2367299
KEYWORDS
SOURCE Escherichia coli K12.
ORGANISM Escherichia coli K12.
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE
AUTHORS 1 (bases 1 to 11509)
Blattner,F.R., Plunkett,G. III, Bloch,C.A., Perna,N.T., Burland,V., Riley,M., Collado-Vides,J., Glasner,J.D., Rode,C.K., Mayhew,G.F., Gregor,J., Davis,N.W., Kirkpatrick,H.A., Goeden,M.A., Rose,D.J., Mau,B. and Shao,Y.
JOURNAL The complete genome sequence of Escherichia coli K-12
MEDLINE Science 277 (331), 1453-1474 (1997)
PUBMED 97426617
REFERENCE 2 (bases 1 to 11509)
AUTHORS Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
3 (bases 1 to 11509)
Blattner,F.R.
TITLE Direct Submission
JOURNAL Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax: 608-263-7459
4 (bases 1 to 11509)
Plunkett,G. III.
TITLE Direct Submission
JOURNAL Submitted (13-OCT-1998) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
On Sep 9, 1997 this sequence version replaced gi:1790254.
This sequence was determined by the E. coli Genome Project at the University of Wisconsin-Madison (Frederick R. Blattner, director). Supported by NIH grants HG00301 and HG01428 (from the Human Genome Project and NCHGR). The entire sequence was independently determined from E. coli K12 strain MG1655. Predicted open reading frames were determined using GenMark software, kindly supplied by Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA, 30332 (e-mail: markambr@gatech.edu). Open reading frames that have been correlated with genetic loci are being annotated with CG Site Nos., unique ID nos. for the genes in the E. coli Genetic Stock Center (CGSC) database at Yale University, kindly supplied by Mary Berlyn. A public version of the database is accessible (<http://cgsc.biology.yale.edu>). Annotation of the genome is an ongoing task whose goal is to make the genome sequence more useful by correlating it with other data. Comments to the authors are appreciated. Updated information will be available at the E. coli Genome Project's World Wide Web site (<http://www.genetics.wisc.edu>). *** The E. coli K12 sequence and its annotations are periodically updated; this is version M54. No sequence changes! Annotation updates: updated gene identifications and products; all new functional assignments courtesy of Monica Riley; added promoters, protein binding sites, and repeated sequences described in reference 1. The unique numeric identifiers beginning with a lowercase 'b' assigned to each gene (protein- or RNA-encoding) are now designated as gene synonyms instead of labels. This should allow them to be searched for in Entrez as gene names.

FEATURES
source
Location/Qualifiers
1..11509
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
81..950
/gene="plda"
/note="b3821"
81..950
/gene="plda"
/FC_number="3.1.1.32"
/function="enzyme; Central intermediary metabolism: phosphorus compounds"
/note="0289; 99 pct identical to PAL_ECOLI SW: P00631; CGSCNO. 384"
/codon_start=1
/transl_table=11
/product="outer membrane phospholipase A"
/protein_id="AAC76824.1"
/db_xref="GI:2367300"
/translation="MRTLGWLLPFLPMAYVQAEATVKEVHADAPAVRGSIANNLQEHNDPPTLYPDYDNLILYQTSIDINKEAIASYDAMENARKDEKFEQLSLAPLMRGILGPSVYLGASTQKSMQOLSNSESSPFRETNEPOLFLGAFIDYRPAQTLDVDEGYNHDSNGRSDPTSRSMRLYTRILMAENGMLVEKPPYVVGNDNDNDPKRYGYVQIKICVHLDGDAVLSAKGYINMNTGIGAGELGISYPTTKHRLTYOVYSGYSGESLDIYNFNOIRVGVVMLNDLF"
965..993
/note="factor sigma70; predicted +1 start at 4003392"
1019..1047
/note="factor sigma70; predicted +1 start at 4003446"
1077..2909
/gene="recQ"
/note="b3822"
1077..2909
/gene="recQ"
/FC_number="3.6.1.-"
/function="enzyme; DNA - replication, repair, restriction/modification"
/note="0610; 99 pct identical to 607 amino acids ofRECQ_ECOLI SW: P15043 but has 3 additional N-ter residues"
/codon_start=1
/transl_table=11
/product="ATP-dependent DNA helicase"
/protein_id="AAC76825.1"
/db_xref="GI:2367301"
/translation="MNVAQAEVLNLESGAKQVLOETFGYQFRPGQEEIITDVLSGRDCLVVPTEGGKSLCYOIPALLNGLTIVVYSPLSIMKQVDOLDQANGVAAACLNSTQTRBOQLVMTGCRGQIRLYIAPERLMDNLFELAHNNPVLAVDEAHICSGWGHFREPVALAGQLRRGPRPILPMALATADPTTRDDIYRLGINDPLQIISFDPNRIRYMIMKRRPDLQRLRYOEOBKGSGITICNSRAVEDTPARLOSKGISAAYHAGLENNYRADVDKFEQRDLDQIVATVAVFGMGLNPNVRFVHPFDPINRISYDQETGAGDGLPAEAMFLYDPADMAMLRCLERKPGQODIERHNLNMGAPAEAKQTCRRLILNLYFGEGROECCGNDICLDPKPYQDSPTQAQLSTIGVYNORFGYGVVEYRGANNRIRDYHGDILKYNGCRBKSHEHWYSVROLIHGLVONIAOHSALQLTFAARPVLAESSLDLAVRYALPKAMOKSGFGGNVRKRFKAKRKIKRKSIAQDSNPPIYVFDATLIEMAEQMPITPASEMLSVNGVMKRLERKGFPMALITAHVDGDDEE"
2104..2140
/note="REP (repetitive extragenic palindromic) element; contains 1 REP sequence"
3162..3189
/note="factor sigma70; predicted +1 start at 4005588"
3224..3592
/gene="yigJ"
/note="yigJ"
/gene="b3823"
3224..3592
/gene="yigJ"
/function="orf; unknown"
/note="0122; sequence change shortens and changesN-terminus relative to earlier version (YIGJ_ECOLI SW: P27846)"
/codon_start=1

```
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC76826.1"
/db_xref="GI:2367302"
/translation="MGYOMLRGALKREAVSAPAPQVELAKSGRSFLKLTJNLNPKA
ITVGSVSILFVGNNGVTARNGIFALITVETLAWFTIVASLFLPQMRNGYRLAKW
IDGRGALFAGSGIHILISR"
repeat_region
3612..3645
/feature="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
complement(3654..4070)
/gene="y1gk"
/feature="y1gk"
complement(3654..4070)
CDS
/feature="orf, Unknown"
/note="f138: 100 pct identical to Y1GK_ECOLI SW: P27847"
/codon_start=1
/transl_table=1
/product="orf, hypothetical protein"
/protein_id="AAC76827.1"
/db_xref="GI:1790258"
/translation="MAEVLKMGAAVLIWLGIQWRAGAIDLKSLASQSRHNF
QRAVYVNTNPKSVTLAALFPQIMPOQPOLMXYIVGTTIVVDIIVMIGYATLLAQ
RIALMTKPRQKALNKIFGSLFMLVCAALLASAKHA"
promoter
4200..4230
/note="factor Sigma70; predicted +1 start at 4006629"
complement(4227..4254)
/promoter="factor Sigma70; predicted +1 start at 4006612"
4384..5406
/gene="p1db"
/note="b3825"
4384..5406
/gene="p1db"
/EC_number="3.1.1.5"
/function="enzyme: Central intermediary metabolism:
phosphorus compounds"
/note="o340: 99 pct identical amino acid sequence and
equal length to P1DB_ECOLI SW: P07000"
/codon_start=1
/transl_table=1
/product="lysophospholipase L(2)"
/protein_id="AAC76828.1"
/db_xref="GI:2367303"
/translation="MFOQKDMETRENAFAFTMGPLTFDEWRQDEAFTGVDDIPVR
FYRFAOHDDRVVVICPGRISVYKAYELADLEHLDVLIIDHRCGRGRLADP
HGHVNRFDNYVDLAFWQOEVOGPGPKRYILLASHSGAISLTLFQIRHFGVCDATA
LTPMFGIVIRKPSFMAKQILLMAEHRPRDGLIGGRKRALPPALVITHSRQK
RNLRFVADDPYIRVGGPTVHWRESILAGEOVLAGAGDDATPTLLQAEERVDNR
MHDRCCELRTAAGHPVEGGRPLVIKGAYHEILFKDAMASVALHAIVDFNRHNSPSG
NRSTEV"
promoter
5381..5408
/note="factor Sigma70; predicted +1 start at 4007807"
5426..5455
/note="factor Sigma70; predicted +1 start at 4007854"
5438..5466
/note="factor Sigma70; predicted +1 start at 4007865"
5494..5523
/note="factor Sigma70; predicted +1 start at 4007922"
5526..6041
/gene="y1gl"
gene
alignment_scores:
Quality: 1024.00 Length: 206
Ratio: 5.020 Gaps: 1
Percent Similarity: 99.029 Percent Identity: 98.058
alignment_block:
US-09-466-935-2 x AE000458/rev ..
Align seg 1/1 to reverse of: AE000458 from: 1 to: 11509
1 MetHrLeuGluTrrTpPheAlaTyrLeuLeuThrSerIleLeuTh 17
```

```
|||||
4273 ATGACCTTAGATGTGGTTGCTTACCTGCTGACATCATGATCTTTAG 4224
17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnH 34
|||||
4223 CCGTGGCCAGGCTGTGGTGCATCAACATCATGACCACTCGCTCAACC 4174
34 lsgLYTrrProAla..GlyGlyValTyrCysTrrPAlaSerAsprrgThg 50
|||||
4173 ACGGTTATCGGCGCGCGCGTGGCTGTATGCTGGCTGACACCGGACTG 4124
50 lYAspSerTrrCysAlaGlyTrrPrrgGlyValGlyThrLeuPheSerArg 66
|||||
4123 GCGATTCTATTGCTGTGCTGCTGGCTGGGGTTGGAGCCTATTTTCCCG 4074
67 SerValIleAlaPheGluValLeuLysTrrPAlaGlyAlaAlaTrrLeuI 83
4073 TCAGTATTCGGTTTGAAGGTGTGAAGTGGCAGCGCGGCTTACTTGAT 4024
83 eTrrLeuGlyllegInGlnTrrPrrgAlaAlaGlyAlaIleAspleuYss 100
4023 TTGGCTGGGAATCCAGCAGTGGCGCGCGCTGGTGCATTGACCTTAAT 3974
100 eTrrLeuAlaSerThrcGlnSerArgrrgHisLeuPheGlnArgAlaValPhe 116
3973 CGCTGGCTTACTCATTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCT 3924
117 ValAsnLeuThrAsnProLysSerIleValPheLeuAlaAlaLeuPrr 133
3923 GCGATTCTACCAATCCCAAGATGTGTGTTGGCGGCGCTATTGCC 3874
133 oGlnPheIleMetProGlnGlnProGlnLeuMetGlnTrrIleValLeuG 150
3873 GCAAATTCATCATGCCGCCAAGCCGCAACTGATGCATATTCGTCGCG 3824
150 lYValThrThrIleValAlaAspIleIleValMetIleGlyTrrAlaThr 166
3823 GCGTACCACTATTGTGGTGCATATTATGATGATGATGATGATGATGATG 3774
167 LeuAlaGlnArgIleAlaLeuTrrPrrleLysGlyProLysGlnMetLysAl 183
3773 CTTGCTCAACGATGCTGCTATGATGATTAAGGACCAAGCAGATGAAGGC 3724
183 aLeuAsnLysIlePheGlySerLeuPheMetLeuValAlaLeuLeuAla 200
3723 GCGTATTAAGATTTTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 3674
200 lAsrAlaArgHisAla 205
3673 CATGGCGGAGGATCGC 3657
seq_name: gb_bal:AE005613
seq_documentation_block:
LOCUS AE005613 10592 bp DNA BCT 25-JAN-2001
DEFINITION Escherichia coli O157:H7 genome, contig 3 of 3, section 232 of 290.
ACCESSION AE005613 AE0005174
VERSION AE005613.1 GI:12518685
KEYWORDS
SOURCE
ORGANISM
Escherichia coli O157:H7.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
REFERENCE
1 (bases 1 to 10592)
Perna,N.T., Plunkett,G., III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L.,
Grothbeck,E.U., Davis,N.W., Lim,A., Dimantanta,B., Potamoustis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
TITLE
JOURNAL
REFERENCE
2 (bases 1 to 10592)
```

AUTHORS Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Haybew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobbeck,E.J., Davis,N.W., Lin,A., Dimalanta,E., Potamousis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.

TITLE Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA

JOURNAL

FEATURES

source 1..10592
Location/Qualifiers
/organism="Escherichia coli O157:H7"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:83334"
/note="enterohemorrhagic"

misc_feature 1..71
/note="O-island #155; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
complement(72..836)
/gene="25332"
complement(72..836)
/gene="25332"
/function="orf; Unknown function"
/note="Residues 1 to 131 of 254 are 95.41 pct identical to residues 1 to 131 of 161 from Escherichia coli K-12 strain MG1655: B3815"
/codon_start=1
/transl_table=11
/protein_id="AAG59008.1"
/db_xref="GI:12518686"
/translation="MAHRLIGKGMITLNRKIFLATLLPLFAVAADDCALSDPTLTVOAYVNPOTERVKMYMOKANGAEAMGTTHALLAMNSOGVOMANGGITYDESAPIGLYEYGQOKVALNLNLSGEGNEFIRPGCVFAGKVCIGVIRDAKTKSEIQFANQSGPMLMENVINPRRIHPNVASRKIRNGVGIKKHGNAYFLISQATNFYDFACYAKALNVEQLYIDGITSIMYMGALIPWORYPVTMISVERKG"

gene 1206..2156
/gene="corA"
/note="25333"
1206..2156
/gene="corA"
/function="transport; Transport of small molecules: cations"
/note="Residues 1 to 316 of 316 are 100.00 pct identical to residues 1 to 316 of 316 from Escherichia coli K-12 strain MG1655: B3816"
/codon_start=1
/transl_table=11
/product="Wg2+ transport, system I"
/protein_id="AAG59009.1"
/db_xref="GI:12518687"
/translation="WLSAFQLENNRRLTRLEVESQPLVNAWVIDIVEPDDDEFRLVOS ELGOSIATREPLEDIEASARPEDDDLHIHSPPFEDADHAGNSVATITDGRIFTLRRELPAFLYRKARASQSMVGNAYELLDFETKIQIOLADEIENTYSIDLEQSLRVIMSGHQDEYDEALSTLALEDIGWVRLCLMDTORALNVLVAKARLPGQLEQARE LIRQESLIPNESLFOKVNFLMQAVGFNIIEQNRIKIKISYVSVFLEPPLVASSY GMEFEMPELKSFGYPGAILFMILAGLAPLYFKRKMWL"

misc_feature 2157..4848
/note="O-island #156; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
complement(2198..2689)
/gene="25334"
complement(2198..2689)
/gene="25334"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59010.1"
/db_xref="GI:12518688"
/translation="MRSKLOKIIICLFICCIYNLMTRPVQVLIITYSDAGNSVFLVY DHEPMTSDKINWTKLHONETKQNDHPPESSHTWIVYIDIGNSFTDYKRIEYEDL YCFPTKSNDCITKNYLVINEYRNTHIGINDTEYQLTQENKIEQVFNHNPKK

gene complement(2686..3534)
/gene="25335"
complement(2686..3534)
/gene="25335"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59011.1"
/db_xref="GI:12518689"
/translation="MNNLLPSSINIPCTIFETISLPDNYSAADMOYGDVVEDEFLSGLSDISAKVDYPRLYIKYHPGPGSINVAFSSSGCTISQRECDIIFAAKELAKFES FPGQVKTILEDLIEHRRYNGSNFHSQQLNSHEKINKYGNSPIRITKECIENGIN STSTGQPLIIDSITKILSLRNKFNDEPSFNLGIVSHDISOKTISLSFOXYA IGWSATIHFAODHFGDLVDYDINKKYSKREFRIFLQRRKDFAKPFPFNPIE RIENYL"

gene 3748..3867
/gene="25336"
3748..3867
/gene="25336"
/function="orf; Unknown function"
/note="Residues 1 to 39 of 39 are 100.00 pct identical to residues 278 to 316 of 316 from GenPept 118 : g116960250|db|AAF33440.1| (AF233324) S. typhimurium magnesium transport protein (COR4) (SP:J31136) [Salmonella typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="partial repeat of corA"
/protein_id="AAG59012.1"
/db_xref="GI:12518690"
/translation="MNFEMPELKSFGYPGAILFMILAGLAPLYFKRKMWL"

gene complement(3892..4278)
/gene="25337"
complement(3892..4278)
/gene="25337"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59013.1"
/db_xref="GI:12518691"
/translation="MEKSSIRNIFPYSKLERLSRPROFIYTLTSMIITINDIWL SFLIVNNYVNIINGAKRIPTGCKAIYFVLISGLHFLIISAIEGVRCESLSVQ LVAVMILFTTILMSAIPSAKMSLSAH"

gene 4483..4794
/gene="25339"
4483..4794
/gene="25339"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59014.1"
/db_xref="GI:12518692"
/translation="MDSIEMWTYPPSCSMYKNPKSKSDPYKAWLIMMAITIVLPI L SITTTUNLSIGISLFGFCQTLGRTLSLGSKIWGVLGGPVTNIVLFLILFFKR R"

gene complement(4841..5728)
/gene="rard"
complement(4841..5728)
/gene="rard"
/function="orf; Unknown function"
/note="Residues 4 to 295 of 295 are 92.12 pct identical to residues 8 to 299 of 300 from Escherichia coli K-12 strain MG1655: B3819"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59015.1"
/db_xref="GI:12518693"

CDS


```
*      16377      16664: gap of unknown length
*      16665      17631: contig of 288 bp in length
*      17632      19241: gap of unknown length
*      19242      20453: contig of 967 bp in length
*      20454      20559: gap of unknown length
*      20560      21117: contig of 1610 bp in length
*      21118      21655: gap of unknown length
*      21656      21818: contig of 1212 bp in length
*      21819      22679: gap of unknown length
*      22680      23579: contig of 163 bp in length
*      23580      24383: gap of unknown length
*      24384      25612: contig of 861 bp in length
*      25613      25957: gap of unknown length
*      25958      27260: contig of 804 bp in length
*      27261      28596: gap of unknown length
*      28597      29473: contig of 345 bp in length
*      29474      30372: gap of unknown length
*      30373      31202: contig of 1303 bp in length
*      31203      32688: gap of unknown length
*      32689      33431: contig of 899 bp in length
*      33432      33693: gap of unknown length
*      33694      34616: contig of 830 bp in length
*      34617      35884: gap of unknown length
*      35885      36459: contig of 923 bp in length
*      36460      36996: gap of unknown length
*      36997      37941: contig of 575 bp in length
*      37942      38682: gap of unknown length
*      38683      39549: contig of 741 bp in length
*      39550      40792: gap of unknown length
*      40793      42585: contig of 867 bp in length
*      42586      44115: gap of unknown length
*      44116      45300: contig of 1243 bp in length
*      45301      46417: gap of unknown length
*      46418      47815: contig of 1185 bp in length
*      47816      49310: gap of unknown length
*      49311      50304: contig of 1398 bp in length
*      50304: gap of unknown length
```

```
*      50305      50848: gap of unknown length
*      50849      51828: contig of 544 bp in length
*      51829      52699: gap of unknown length
*      52700      53365: gap of unknown length
*      53366      54633: contig of 871 bp in length
*      54634      55471: gap of unknown length
*      55471      56359: contig of 666 bp in length
*      56359      57672: gap of unknown length
*      57672      58319: contig of 1268 bp in length
*      58320      58529: gap of unknown length
*      58530      59083: contig of 837 bp in length
*      59084      59906: gap of unknown length
*      59907      60826: contig of 888 bp in length
*      60827      61084: gap of unknown length
*      61085      61762: contig of 576 bp in length
*      61763      62639: gap of unknown length
*      62639      62714: contig of 738 bp in length
*      62714      63594: gap of unknown length
*      63594      64039: contig of 876 bp in length
*      64040      65077: gap of unknown length
*      65077      65367: contig of 75 bp in length
*      65367      65591: gap of unknown length
*      65591      66041: contig of 446 bp in length
*      66042      67040: gap of unknown length
*      67041      67560: contig of 451 bp in length
*      67560: gap of unknown length
```

alignment_scores: Quality: 795.00 Length: 174
 Ratio: 4.789 Gaps: 3
Percent Similarity: 95.402 Percent Identity: 93.678

alignment_block:
US-09-466-935-2 x AC020870/rev ..

Align seg 1/1 to reverse of: AC020870 from: 1 to: 256373

```
1 Metrrlauglurtrpphealatyrlaurhserrllaleuth 17
|||||
77559 ATGACCTTAGAATGCTGCTTGCCTACCTGACATCATTTTAA 77510
|||||
17 rleu..Serproglysergilalileasphmethrthserleas 33
|||||
77509 CCTGCTGCCCAAGGCTCTGTGCATCAACATATGACCACTGCTCAA 77460
|||||
33 nhtslgtyrprola..glylyvaltyrcysrpalaseraaryth 49
|||||
```


77459 CCAGGTTATCGGCGCGGTCGTCATTCCTGGGCTTCAGACCGGAC 77410
49 rGlyAspSerTyrCysAlaGlyTrpArgGlyVal GlyThrLeuPheSer 65
|||||
77409 TGGCGATTCAATTGTCGTGGTGGGCTGGGGGACGCTATTATTTCC 77360
66 ArgSerValIleAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLe 82
|||||
77359 CGCTCAGTGAATGCGTTGAAGTGTGAAGTGGCAGCGCGCTTACTT 77310
82 uileTrpLeuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspLeu 99
|||||
77309 GATTGGGTGGAAATCCAGCATGGCGCGCCCTGTGTCATATGACCTTA 77260
99 ySSerLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaVal 115
|||||
77259 AATCGCTGCGCTCTACTCATCCGTCGACATTTGTCACGCGCAGTT 77210
116 PheValAsnLeuPheAsnProLysSerIleValPheLeuAlaIleLeuPhe 132
|||||
77209 TTTGTGATCTCCACCAATCCCAAAAGTATTGTCTTGGCGCGCTATT 77160
132 eProGlnPheIleMetProGlnGlnProGlnLeuMetGlnTyrIleVal 149
|||||
77159 TCCCAATTATCATCATGCCGACAGCCGCACTGATGAGATATCGTGC 77110
149 euGlyValIThrThrIleValIleValAspIleIleValMetIleGlyTyrAla 165
|||||
77109 TCGGCGTCACCACTATTGTGTCGATATATTGTGATGATGATGATGATG 77060
166 ThrLeuAlaGlnArgIle 171
|||||
77059 ACCTGCTCTACACAGGAGTC 77042
seq_name: gb_hhg8:AC022157
seq_documentation_block: 4833 bp DNA HTG 26-JAN-2000
LOCUS AC022157 Homo sapiens chromosome 5 clone RPI-280K18, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
ACCESSION AC022157
VERSION AC022157.1 GI:6758618
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4833)
AUTHORS DOE Joint Genome Institute.
TITLE Unpublished
JOURNAL 2 (bases 1 to 4833)
REFERENCE DOE Joint Genome Institute.
AUTHORS Direct Submission
TITLE Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint
JOURNAL Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94558, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
* NOTE: This record contains 8 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1 836: contig of 836 bp in length

* * * gap of unknown length
* 837 1566: contig of 730 bp in length
* * * gap of unknown length
* 1567 2612: contig of 1046 bp in length
* * * gap of unknown length
* 2613 2721: contig of 109 bp in length
* * * gap of unknown length
* 2722 2829: contig of 108 bp in length
* * * gap of unknown length
* 2830 3052: contig of 223 bp in length
* * * gap of unknown length
* 3053 4256: contig of 1204 bp in length
* * * gap of unknown length
* 4257 4833: contig of 577 bp in length.
FEATURES
source
1. 4833
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="5"
/clone="RPI-280K18"
BASE COUNT 1218 a 1298 c 1250 g 1055 t 12 others
ORIGIN
alignment_scores:
Quality: 751.50 Length: 211
Ratio: 4.175 Gaps: 5
Percent Similarity: 85.308 Percent Identity: 77.251
alignment_block:
US-09-466-935-2 x AC022157/rev ..
Align seq 1/1 to reverse of: AC022157 from: 1 to: 4833
10 LeuLeuThrSerIleIleLeuThrLeuSerProGlySerGlyAlaIleAs 26
|||||
878 ATTGTGCTGCGCATTCCTCTATCTCATCGCTCAGCTCTGCTGC 840
26 rThrMetThrThrSerLeuAsnHisGlyTyrProAla..GlyGlyValTyr 42
|||||
839 AGGATGACCACTCGCTCAGACACGCGTATGCGCGCGGCGGCTCTTA 791
42 rCySTrpAlaSerAspArgThrGlyAspSerTyrCysAlaGlyTrpArg. 58
|||||
790 TTGCTGGGCTTCAGACCGGACGCGCATTCATATTGCTGTTGGCGTG 741
59 GlyValGlyThrLeuPheSerArgSerValIleAlaPheGluValLeuLys 75
|||||
740 GCGTTGGGTACGCTATTTCGCCGTCAGATGCGGCTTGAAGTGTGAA 691
75 sTrpAlaGlyAlaAlaTyrLeuIleTrpLeuGlyIleGlnGlnTrpArg 92
|||||
690 GTGGCGGCGCGCGCTTACTGATTTGGCTGGGATTCACAGAGCGCGCG 641
92 IaAlaGlyAlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgArg 108
|||||
640 CCGTGTGTCGAATTCATTAAATCGCTGCGCTTACTCATTCGCGTGA 591
109 HisLeuPheGlnArgAlaValPheValAsnLeuThrAsnProLysSerI 125
|||||
590 CATTTGTTCCAGCGCGCAGATTTTGGATTCACCAATCCCAAAAGTAT 541
125 eValPheLeuAlaIleLeuPheProGlnPheIleMetProGlnGlnProG 142
|||||
540 TGTGTTTCTGGCGCGCTATTTCGCAATTCAATTCATGCGGACAGCGCG 491
142 InLeuMetGlnTyr..... 146
490 AACGATGACAGATTAGTCGTGCGCGCGGTCGAGCGCAGCGGTAGGT 441
147 ...IleValLeuGlyValThrThrIleValIleValAspIleIleValMet 162
|||||
440 GTGGCTGTGTCGAGTATGATGATTTTC.GTCGAGTGAATGCGGCTTA.. 394

162 eGIYrAlaThrlaAlaGlnArgIleAlaLeuTrpIleGlyProl 179
|||||.....
393CCGGACGCGCTTGCTCAACGATGCTCTATGATTAAGACCAA 348
179 ySGInMetLysAlaLeuAsnLysIlePheGlySerLeuPheMetLeuVal 195
|||||.....
347 AGCAGATGAAGCGCGCTGAATAGATTTCGCTGTTTATGCTGTG 298
196 GLyAlaLeuLeuAlaSerAlaArgHisAla 205
|||||.....
297 GGAGCGCTGTACATCGCGCAGGCATCCG 268

seq_name: gb_hcg7:AC020970

seq_documentation_block:
LOCUS AC020970 212936 bp DNA HTG 10-FEB-2000
DEFINITION Mus musculus clone Rp23-252w21, WORKING DRAFT SEQUENCE, 144
unordered pieces.
ACCESSION AC020970
VERSION AC020970.1 GI:6691260
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
1 (bases 1 to 212936)
DOE Joint Genome Institute.
Unpublished
2 (bases 1 to 212936)
DOE Joint Genome Institute.
Direct Submission
Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

-----Summary Statistics
Estimated insert size: 212936; sum-of-contigs estimation
Estimated insert size: 157000; pulse field gel estimation
Quality coverage: 4.20x in Q20 bases; pulse field gel estimation
Quality coverage: 3.10x in Q20 bases; sum-of-contigs estimation

NOTE: This is a 'working draft' sequence. It currently
* consists of 144 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
1 1159: contig of 1159 bp in length
* gap of unknown length
*
* 1160 2363: contig of 1204 bp in length
* gap of unknown length
*
* 2364 3604: contig of 1241 bp in length
* gap of unknown length
*
* 3605 5188: contig of 1584 bp in length
* gap of unknown length
*
* 5189 6372: contig of 1184 bp in length
* gap of unknown length
*
* 6373 7737: contig of 1365 bp in length
* gap of unknown length
*
* 7738 9783: contig of 2046 bp in length
* gap of unknown length
*
* 9784 11046: contig of 1263 bp in length
* gap of unknown length
*
* 11047 12223: contig of 1177 bp in length
* gap of unknown length
*
* 12224 13492: contig of 1269 bp in length
* gap of unknown length

* 13493 14669: contig of 1177 bp in length
* gap of unknown length
*
* 14670 15886: contig of 1217 bp in length
* gap of unknown length
*
* 15887 17089: contig of 1203 bp in length
* gap of unknown length
*
* 17090 18308: contig of 1219 bp in length
* gap of unknown length
*
* 18309 19518: contig of 1210 bp in length
* gap of unknown length
*
* 19519 20710: contig of 1192 bp in length
* gap of unknown length
*
* 20711 21932: contig of 1222 bp in length
* gap of unknown length
*
* 21933 23606: contig of 1674 bp in length
* gap of unknown length
*
* 23607 25328: contig of 1722 bp in length
* gap of unknown length
*
* 25329 26649: contig of 1321 bp in length
* gap of unknown length
*
* 26650 27849: contig of 1200 bp in length
* gap of unknown length
*
* 27850 29135: contig of 1286 bp in length
* gap of unknown length
*
* 29136 30700: contig of 1563 bp in length
* gap of unknown length
*
* 30701 31877: contig of 1177 bp in length
* gap of unknown length
*
* 31878 33065: contig of 1188 bp in length
* gap of unknown length
*
* 33066 34267: contig of 1202 bp in length
* gap of unknown length
*
* 34268 35474: contig of 1207 bp in length
* gap of unknown length
*
* 35475 36650: contig of 1176 bp in length
* gap of unknown length
*
* 36651 37859: contig of 1209 bp in length
* gap of unknown length
*
* 37860 39072: contig of 1213 bp in length
* gap of unknown length
*
* 39073 40406: contig of 1334 bp in length
* gap of unknown length
*
* 40407 41630: contig of 1224 bp in length
* gap of unknown length
*
* 41631 42864: contig of 1234 bp in length
* gap of unknown length
*
* 42865 44200: contig of 1336 bp in length
* gap of unknown length
*
* 44201 46087: contig of 1887 bp in length
* gap of unknown length
*
* 46088 48227: contig of 2140 bp in length
* gap of unknown length
*
* 48228 49518: contig of 1291 bp in length
* gap of unknown length
*
* 49519 50910: contig of 1392 bp in length
* gap of unknown length
*
* 50911 52420: contig of 1510 bp in length
* gap of unknown length
*
* 52421 53613: contig of 1193 bp in length
* gap of unknown length
*
* 53614 55004: contig of 1391 bp in length
* gap of unknown length
*
* 55005 56300: contig of 1296 bp in length
* gap of unknown length
*
* 56301 57505: contig of 1205 bp in length
* gap of unknown length
*
* 57506 58686: contig of 1181 bp in length
* gap of unknown length
*
* 58687 60407: contig of 1721 bp in length
* gap of unknown length
*
* 60408 61632: contig of 1225 bp in length
* gap of unknown length
*
* 61633 62880: contig of 1248 bp in length

*		gap of unknown length
*	62881	64220: contig of 1340 bp in length
*		gap of unknown length
*	64221	65420: contig of 1200 bp in length
*		gap of unknown length
*	65421	66605: contig of 1185 bp in length
*		gap of unknown length
*	66606	68682: contig of 2077 bp in length
*		gap of unknown length
*	68683	70712: contig of 2030 bp in length
*		gap of unknown length
*	70713	71895: contig of 1183 bp in length
*		gap of unknown length
*	71896	73462: contig of 1567 bp in length
*		gap of unknown length
*	73463	74806: contig of 1344 bp in length
*		gap of unknown length
*	74807	76205: contig of 1399 bp in length
*		gap of unknown length
*	76206	77580: contig of 1375 bp in length
*		gap of unknown length
*	77581	79081: contig of 1501 bp in length
*		gap of unknown length
*	79082	80314: contig of 1233 bp in length
*		gap of unknown length
*	80315	81899: contig of 1585 bp in length
*		gap of unknown length
*	81900	83140: contig of 1241 bp in length
*		gap of unknown length
*	83141	84665: contig of 1525 bp in length
*		gap of unknown length
*	84666	85888: contig of 1223 bp in length
*		gap of unknown length
*	85889	87155: contig of 1267 bp in length
*		gap of unknown length
*	87156	88435: contig of 1280 bp in length
*		gap of unknown length
*	88436	89955: contig of 1520 bp in length
*		gap of unknown length
*	89956	91277: contig of 1322 bp in length
*		gap of unknown length
*	91278	92513: contig of 1236 bp in length
*		gap of unknown length
*	92514	93917: contig of 1404 bp in length
*		gap of unknown length
*	93918	95448: contig of 1531 bp in length
*		gap of unknown length
*	95449	96800: contig of 1352 bp in length
*		gap of unknown length
*	96801	97978: contig of 1178 bp in length
*		gap of unknown length
*	97979	99179: contig of 1201 bp in length
*		gap of unknown length
*	99180	100437: contig of 1258 bp in length
*		gap of unknown length
*	100438	101827: contig of 1390 bp in length
*		gap of unknown length
*	101828	103117: contig of 1290 bp in length
*		gap of unknown length
*	103118	104550: contig of 1433 bp in length
*		gap of unknown length
*	104551	105854: contig of 1304 bp in length
*		gap of unknown length
*	105855	107523: contig of 1669 bp in length
*		gap of unknown length
*	107524	109208: contig of 1685 bp in length
*		gap of unknown length
*	109209	110430: contig of 1222 bp in length
*		gap of unknown length
*	110431	111662: contig of 1232 bp in length

alignment_scores:
Quality: 734.00

Length: 206

	Ratio: 3.745	Gaps: 1
	Percent Similarity: 95.146	Percent Identity: 93.204
	alignment_block:	
	US-09-466-935-2 x AC020970/rev ..	
	Align seg 1/1 to reverse of: AC020970 from: 1 to: 212936	
1	MetThrLeuGluThrPrpPheAlaTyrLeuLeuThrSerIleLeuTh	17
175539	ATGACCTTGAAATGATGGTTGGCTACCTGCTGCA.TGCATCATTTTAA	175491
17	rLeuSerProGlySerGlyAlaIleasnThrMetThrSerLeuAsn	34
175490	CCGTGGCCAGGCTGTGTCATCAACACTAT.ACCACCTGCTCAACC	175442
34	IsGlyTyrProAla..GlyGlyValTyrCysThrPheAlaSerArgTh	50
175441	ACCGTTATGCGGCGCGGCTGTATGCTGGCT.TCAGACCGGACTG	175393
50	LYASPERTYRCYSALAGLYTPRARGLYVALGlyThrLeuPheSerArg	66
175392	GCGATTCAATATGTGCTGTTGGCGTGGGCTTGGAGCGCTATTTCGCC	175343
67	SerValIleAlaPheGluValLeuIstYrPheAlaGlyAlaIleTyrLeu	83
175342	TCAGTATTCGCTTGAACTGTAAGTGGGC.GGCGCGCTTACTTGAT	175294
83	eTrrPheGlyIleGlnGlnThrPheAlaIleAlaIleAspLeuLys	100
175293	TTGGCTGGGAATCCACAGTGGCGGCGCTGTGCACATTCACCTTAAT	175244
100	eTrrPheAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValPhe	116
175243	CGGTGGCCCTCTACTGATGGCTGCACATTT.TTCCAGCGCCAGTTT	175195
117	ValAsnLeuThrAsnProLysSerIleValPheLeuAlaIleLeuPheTr	133
175194	GTGAATCTCACCAATCCAAAGTATTGTCTCG.GCGGCGCTATTGCC	175146
133	oGlnPheIleMetProGlnGlnProGlnLeuMetGlnTyrIleValLeuG	150
175145	GCAATTCATATGCGCACAGCCGCACTGTC.CAGTATATCGTGCCTG	175097
150	LYVALTHRTHRILEVALVALASPILEILEVALMETILEGLTYRALATHR	166
175096	GGGTACCACTATTGTGTCGATATTATTGT.ATGATCGTTACGCCAC	175048
167	LeuAlaGlnArgIleAlaLeuThrPheIleLysGlyProLysGlnMetLysAl	183
175047	CTTGGCCACAGCATTCCTCATGATTAAAGA.CCAAGCAGATGAAGC	174999
183	AlaLeuAsnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeuAla	200
174998	GCTGAATTAAGATTTTGGCTGCTGTATTATGTC.GTGGAGCGCTGTAG	174950
200	LaSerAlaArgHisAla 205	
174949	CATCGCGCAGCATCGC 174933	
seq_name: gb_htg7:AC020874		
seq_documentation_block:		
LOCUS	AC020874 265383 bp DNA	HTG 16-FEB-2000
DEFINITION	Mus musculus clone Rp23-333M14, LOW-PASS SEQUENCE	SAMPLING.
ACCESSION	AC020874	
VERSION	AC020874.2 GI:6980207	
KEYWORDS	HTG; HTGS_PHASE0.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
	1 (bases 1 to 265383)	

AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 265383)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On Feb 16, 2000 this sequence version replaced gi:6686434.
* NOTE: This record contains 244 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 869: contig of 869 bp in length
* *
* 870 1181: contig of 312 bp in length
* *
* 1182 2202: contig of 1021 bp in length
* *
* 2203 2960: contig of 758 bp in length
* *
* 2961 3646: contig of 686 bp in length
* *
* 3647 4520: contig of 874 bp in length
* *
* 4521 4813: contig of 293 bp in length
* *
* 4814 5434: contig of 621 bp in length
* *
* 5435 6000: contig of 566 bp in length
* *
* 6001 6992: contig of 992 bp in length
* *
* 6993 7785: contig of 793 bp in length
* *
* 7786 8367: contig of 582 bp in length
* *
* 8368 9114: contig of 747 bp in length
* *
* 9115 9729: contig of 615 bp in length
* *
* 9730 10313: contig of 584 bp in length
* *
* 10314 10712: contig of 399 bp in length
* *
* 10713 11404: contig of 692 bp in length
* *
* 11405 12261: contig of 857 bp in length
* *
* 12262 13173: contig of 912 bp in length
* *
* 13174 13934: contig of 761 bp in length
* *
* 13935 14703: contig of 769 bp in length
* *
* 14704 15380: contig of 677 bp in length
* *
* 15381 15654: contig of 274 bp in length
* *
* 15655 16714: contig of 1060 bp in length
* *
* 16715 17249: contig of 535 bp in length
* *
* 17250 17863: contig of 614 bp in length
* *
* 17864 18508: contig of 645 bp in length

* 18509 19193: gap of unknown length
* 19194 19737: gap of unknown length
* 19738 20600: gap of unknown length
* 20601 21239: gap of unknown length
* 21240 21903: gap of unknown length
* 21904 22705: gap of unknown length
* 22706 22948: gap of unknown length
* 22949 23682: gap of unknown length
* 23683 24364: gap of unknown length
* 24365 25288: gap of unknown length
* 25289 25954: gap of unknown length
* 25955 26094: gap of unknown length
* 26095 27015: gap of unknown length
* 27016 27740: gap of unknown length
* 27741 28060: gap of unknown length
* 28061 28343: gap of unknown length
* 28344 29016: gap of unknown length
* 29017 29371: gap of unknown length
* 29372 29586: gap of unknown length
* 29587 29691: gap of unknown length
* 29692 30501: gap of unknown length
* 30502 31132: gap of unknown length
* 31133 31249: gap of unknown length
* 31250 32467: gap of unknown length
* 32468 32878: gap of unknown length
* 32879 34069: gap of unknown length
* 34070 34694: gap of unknown length
* 34695 35933: gap of unknown length
* 35934 36920: gap of unknown length
* 36921 37173: gap of unknown length
* 37174 37282: gap of unknown length
* 37283 38196: gap of unknown length
* 38197 38452: gap of unknown length
* 38453 38678: gap of unknown length
* 38679 38835: gap of unknown length
* 38836 39561: gap of unknown length


```
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
*
*      1      635: contig of 635 bp in length
*      *      gap of unknown length
*      *      636      1342: contig of 707 bp in length
*      *      gap of unknown length
*      *      1343      1956: contig of 614 bp in length
*      *      gap of unknown length
*      *      1957      2656: contig of 700 bp in length
*      *      gap of unknown length
*      *      2657      2837: contig of 181 bp in length
*      *      gap of unknown length
*      *      2838      3485: contig of 648 bp in length
*      *      gap of unknown length
*      *      3486      4071: contig of 586 bp in length
*      *      gap of unknown length
*      *      4072      4571: contig of 500 bp in length
*      *      gap of unknown length
*      *      4572      4875: contig of 304 bp in length
*      *      gap of unknown length
*      *      4876      5560: contig of 685 bp in length
*      *      gap of unknown length
*      *      5561      6211: contig of 651 bp in length
*      *      gap of unknown length
*      *      6212      6915: contig of 704 bp in length
*      *      gap of unknown length
*      *      6916      7565: contig of 650 bp in length
*      *      gap of unknown length
*      *      7566      7879: contig of 314 bp in length
*      *      gap of unknown length
*      *      7880      8054: contig of 175 bp in length
*      *      gap of unknown length
*      *      8055      9082: contig of 1028 bp in length
*      *      gap of unknown length
*      *      9083      9199: contig of 117 bp in length
*      *      gap of unknown length
*      *      9200      10135: contig of 936 bp in length
*      *      gap of unknown length
*      *      10136      10450: contig of 315 bp in length
*      *      gap of unknown length
*      *      10451      11268: contig of 818 bp in length
*      *      gap of unknown length
*      *      11269      12291: contig of 1023 bp in length
*      *      gap of unknown length
*      *      12292      12948: contig of 657 bp in length
*      *      gap of unknown length
*      *      12949      13627: contig of 679 bp in length
*      *      gap of unknown length
*      *      13628      14182: contig of 555 bp in length
*      *      gap of unknown length
*      *      14183      15036: contig of 854 bp in length
*      *      gap of unknown length
*      *      15037      15215: contig of 179 bp in length
*      *      gap of unknown length
*      *      15216      15974: contig of 759 bp in length
*      *      gap of unknown length
*      *      15975      16596: contig of 622 bp in length
*      *      gap of unknown length
*      *      16597      16663: contig of 67 bp in length
*      *      gap of unknown length
*      *      16664      16954: contig of 291 bp in length
*      *      gap of unknown length
*      *      16955      17608: contig of 654 bp in length
*      *      gap of unknown length
*      *      17609      17762: contig of 154 bp in length
*      *      gap of unknown length
*      *      17763      17995: contig of 233 bp in length
*      *      gap of unknown length
*      *      17996      18075: contig of 80 bp in length
*      *      gap of unknown length
*      *      18076      18233: contig of 158 bp in length
*      *      gap of unknown length
*
*      18234      18936: contig of 703 bp in length
*      *      gap of unknown length
*      *      18937      19076: contig of 140 bp in length
*      *      gap of unknown length
*      *      19077      19299: contig of 223 bp in length
*      *      gap of unknown length
*      *      19300      19490: contig of 191 bp in length
*      *      gap of unknown length
*      *      19491      19705: contig of 215 bp in length
*      *      gap of unknown length
*      *      19706      19865: contig of 160 bp in length
*      *      gap of unknown length
*      *      19866      20017: contig of 152 bp in length
*      *      gap of unknown length
*      *      20018      20115: contig of 98 bp in length
*      *      gap of unknown length
*      *      20116      21069: contig of 954 bp in length
*      *      gap of unknown length
*      *      21070      22416: contig of 1347 bp in length
*      *      gap of unknown length
*      *      22417      23017: contig of 601 bp in length
*      *      gap of unknown length
*      *      23018      23192: contig of 175 bp in length
*      *      gap of unknown length
*      *      23193      23549: contig of 357 bp in length
*      *      gap of unknown length
*      *      23550      24450: contig of 901 bp in length
*      *      gap of unknown length
*      *      24451      24820: contig of 370 bp in length
*      *      gap of unknown length
*      *      24821      25794: contig of 974 bp in length
*      *      gap of unknown length
*      *      25795      26071: contig of 277 bp in length
*      *      gap of unknown length
*      *      26072      26717: contig of 646 bp in length
*      *      gap of unknown length
*      *      26718      26912: contig of 195 bp in length
*      *      gap of unknown length
*      *      26913      27578: contig of 666 bp in length
*      *      gap of unknown length
*      *      27579      28761: contig of 1183 bp in length
*      *      gap of unknown length
*      *      28762      28931: contig of 170 bp in length
*      *      gap of unknown length
*      *      28932      29143: contig of 212 bp in length
*      *      gap of unknown length
*      *      29144      30237: contig of 1094 bp in length
*      *      gap of unknown length
*      *      30238      30693: contig of 456 bp in length
*      *      gap of unknown length
*      *      30694      31274: contig of 581 bp in length
*      *      gap of unknown length
*      *      31275      31510: contig of 236 bp in length
*      *      gap of unknown length
*      *      31511      31680: contig of 170 bp in length
*      *      gap of unknown length
*      *      31681      31733: contig of 53 bp in length
*      *      gap of unknown length
*      *      31734      32506: contig of 773 bp in length
*      *      gap of unknown length
*      *      32507      33123: contig of 617 bp in length
*      *      gap of unknown length
*      *      33124      33855: contig of 732 bp in length
*      *      gap of unknown length
*      *      33856      34744: contig of 889 bp in length
*      *      gap of unknown length
*      *      34745      35489: contig of 745 bp in length
*      *      gap of unknown length
*      *      35490      35600: contig of 111 bp in length
*      *      gap of unknown length
*      *      35601      36330: contig of 730 bp in length
*      *      gap of unknown length
*      *      36331      36544: contig of 214 bp in length
```


gene
HHRP
2105. . 2722
/gene="YC0136"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF9313.1"
/db_xref="GI:9654537"
/translation="MDITWVLYLITFAVPSLQSGTVNSIGLSYGRHSIGATI
GLQISLAHIYVIGIGISALVQSLAFTLTKIGAYLYVMGIGQKRRSLPATIT
SHEISQALKKRAVLINLTNPKSIIVFVLAEPQIFDIPDRHMPQFVLGITYVTIDT
VMFGTALAAQIGRIYRSPNIMTRMKLFGSMCGMLATAKA"
2670. . 3968
/gene="YC0137"
2670. . 3968
/gene="YC0137"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF9314.1"
/db_xref="GI:9654538"
/translation="MYRCIMAECCMLPRKKRPFEDTWTITTYAROPILNARKHTLG
TELLERREKLSRTQKRALCRINACARHELLPEODATYNNQTLDIRFPADSTIDHKRQ
PLTPRREDAIVEILFETCPQPTDLEAVQELSORGYLLALDDEVSPAMERFLPVQVY
KTIIDKAMLDACEFNRGLAQSGRRFLERVETDEDFHQRHAGTFPOGFFSKP
ELIKORYYSEPHVIMOLFREVCOPEVDYVERVLVADIALSYKLRFVNMDSRS
VSISSFRQAIYVIGODKLRITFVSLAVSYSSKKPELYNLSIQRAQFCOLMATHFRF
KAHRQQAQILFESYDALIDPTSLIQLVQYEOIPLADVDKIALREHSGPLGLLIDEECF
EKADQGVGEQHCLEGLFDEYRQELITPAQRWSDINRLI"
complement(3976. .4095)
/gene="YC0138"
complement(3976. .4095)
/gene="YC0138"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AAF9315.1"
/db_xref="GI:9654539"
/translation="WVGQPLSSYTLRWGLADHLSPNDLPLPTSLPQVEKXH"
complement(4148. .4618)
/gene="YC0139"
complement(4148. .4618)
/gene="YC0139"
/note="similar to GB:IL42023 SP:P45173 PID:1007363
PID:1221482 PID:1205585; identified by sequence
similarity; putative"
/codon_start=1
/transl_table=1
/product="DPS family protein"
/protein_id="AAF9316.1"
/db_xref="GI:9654540"
/translation="MATNLIGDITQSOKLANLNLLANYQVEYNTRGYHWNIOCKR
EPEFLHAKREERYTDLQKIDELAIRIITLISARPHSFGYLAKAOKIENHDSIDGSS
SMQGVLDGSIILHQORDILFELAGTEGGSALMSDYIRQEKLVMLNWLK"
complement(4788. .5117)
/gene="YC0140"
complement(4788. .5117)
/gene="YC0140"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AAF9317.1"
/db_xref="GI:9654541"
/translation="MPFTLLSLALLITIAAVFSSHFRHMGENAPQRSAAEVTIID
KQAIDPHLPQDDQEFWIIYQKRGRLCPKRKEFOVGIIHYHALNPGCKGVALTGQGRFRF
LHFALKR"

```

gene      5156. .5260
          /gene="VC0141"
CDS       5156. .5260
          /gene="VC0141"
          /note="identified by glimmer2; putative"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAFP93318.1"
          /db_xref="GI:9654542"
          /translation="MYFSKQFVRRKCAAVIGIFSAYLTLQPLNG"
          5322. .5690
          /gene="VC0142"
          5332. .5690
          /gene="VC0142"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAFP93319.1"
          /db_xref="GI:9654543"
          /translation="MSEIDQVLRTRKLETLLEQRYAEBKGGLHQLITSGEERPHYV
          IKLRVYATIRNKRVEDDDYRLDDRKAFLVACDEDELPRSGKFWRIAFMLMLM
          TLALPFYVHNRWMBELSOHQ"
          6193. .6450
          /gene="VC0143"
          6193. .6450
          /gene="VC0143"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=11
          /product="hypothetical protein"
          /protein_id="AAFP93320.1"
          /db_xref="GI:9654544"
          /translation="MELSPICFYFRKGSYSPFCIAGCRQDFVVFVFIQGLFLFC
          SEGLVDSEFYFLVYHLFGVDLISFGMKQVPLFCNFI"
          6567. .7196
          /gene="VC0144"
          6567. .7196
          /gene="VC0144"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=11
          /product="conserved hypothetical protein"
          /protein_id="AAFP93321.1"
          /db_xref="GI:9654545"
          /translation="MSMLSMTSVLSGFLSISAVENOKLQAIFFRIISLLILTIYW
          EOHSHATPEVIFETISGLAVSFMAGHLNHRHYKASVPLFLVAQLLSKAWVDSGS
          MYWMLPALIVAASTVAFELLIPQDITLIPFTINGMLNOMTAAAGELMIOEAIVASGS
          AGFLGCVIITISATLLAIHIDRRPLPWGHTLISSYLLAOLIASIVF"
          complement(7245. .7511)
          /gene="VC0145"
          complement(7245. .7511)
          /gene="VC0145"
          /note="identified by Glimmer2; putative"
          /codon_start=1
          /transl_table=11

alignment_scores:
          Quality: 462.50      Length: 206
          Ratio: 2.855      Gaps: 3
Percent Similarity: 78.641      Percent Identity: 44.660

alignment_block:
US-09-466-935-2 x AEO04104 ..

Align seg 1/1 to: AEO04104 from: 1 to: 11305

1 MetThrLeuGluTrpIrrpPheAlaIrrLeuLeuThSerIleLeuTh 17
||||| : : : : : ||| : : : : : ||| : : : : : ||| : : : : :
2105 AAGGATATCCATGTTGGCTGCGCTATCTACGACCGCGGTGTTTAA 2154

17 rIeuSerProGlySerGlyAlaIleAsnThrMetThrTrhSerLeuAsn 34

```


seq_name: gb_ba1:AE004937
seq_documentation_block:
LOCUS AE004937 10475 bp DNA BCT 30-AUG-2000
DEFINITION Pseudomonas aeruginosa PA01, section 498 of the complete genome.
ACCESSION AE004937 AE004091
VERSION AE004937.1 GI:9951553
KEYWORDS
SOURCE Pseudomonas aeruginosa.
ORGANISM Pseudomonas aeruginosa
Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
Pseudomonas
1 (bases 1 to 10475)
Stover,C.K., Pham,X.Q., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S., Hutnagle,W.O., Kowalik,D.J., Lagrou,M.,
Garber,R.L., Goltzy,L., Tolentino,E., Westbrook-Wadman,S., Yuan,Y.,
Brody,L.L., Coulter,S.N., Folger,K.R., Kas,A., Lardig,K., Lim,R.,
Smith,K., Spencer,D., Wong,G.K., Wu,Z. and Paulsen,I.T.
Complete genome sequence of Pseudomonas aeruginosa PA01, an
opportunistic pathogen
Nature 406 (6799), 959-964 (2000)
JOURNAL MEDLINE 20437337
REFERENCE 2 (bases 1 to 10475)
AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J.,

2204 TCTGCGCCCGGCTTCAGCAGCGCTTAACCTGCATGCACCAATGGTTTAAGTT
34 isgly.....TyrProAlaGlyGlyValTyrCysTrpAlaSerAspArg 48
2205 ATGGTACAGCCCACTCATCTAGGGGCGCATATTCGG....CCTACCAATTCG 2250
49 ThrGlyAspSerTyrCysAlaGlyTTPArg.GlyValGlyTThLeuPhe 65
2251 TCTTGCTTGTCATATGCTGTGGTGGCATTTGGAATTTGGCGCTTGGTGG 2300
65 erArSerValIleAlaPheGluValLeuYsrTrpAlaGlyAlaIaTyr 81
2301 CACATATGGCTGGTGGCTTCACCTCATCAATGCATGGCGGCGCTTAT 2350
82 LeuIleTrpLeuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAsp 98
2351 TTAGTATGGCTGGGATTTCAGAAATGGCGAGATAGAGCCCGCTTAACG 2400
98 ulYsrSerLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAla 115
2401 TACCACAACTTCTCATGAACTCTCACAGCGCGCTGCTCGTAAAGCAG 2450
115 alPheValAsnLeuThrAsnProYsrSerIleValPheLeuAlaIaLeu 131
2451 TACTGATTAACCTGACCAATCCCAATCATCGTTTCTTAGTGGCATTA 2500
132 PheProGlnPheIleMetProGlnInProGlnLeuMetGlnTyrIleVal 148
2501 TTTCGGCAGTTTATTTGATCCCACTAGATCATTTGGCGCAGTTCTTGGT 2550
148 lleGlyValThrThrIleValAlaIleIleValMetIleGlyTyr 165
2551 TCTGGGATACCAACCGTTACTATGATGCGATTTGCATCTTCGGATACA 2600
165 IatTrIleAlaGlnArgIleAlaLeuTrpIleLeuGlyPolYsglnMet 181
2601 CGGCACGTGGCGCAGCTAGGTGCTATATCTTACCTAACATATATG 2650
182 LysAlaLeuAsnLysIlePheGlySerLeuPheMetLeuValGlyAla 198
2651 ACCAGAAATGAATAAAGTTTGGTTCGATGTTATGGCTGCGGAATGCT 2700
198 ulLeuAlaSerAlaArg 203
2701 GTTGGCTACCGCAAA 2716
TITLE
JOURNAL MEDLINE 20437337
REFERENCE 2 (bases 1 to 10475)
AUTHORS Stover,C.K., Pham,X.-Q.T., Erwin,A.L., Mizoguchi,S.D., Warrenner,P.,
Hickey,M.J., Brinkman,F.S.L., Hutnagle,W.O., Kowalik,D.J.,

FEATURES
source

Lagrou,M., Garber,R.L., Goltzy,L., Tolentino,E.,
Westbrook-Wadman,S., Yuan,Y., Brody,L.L., Coulter,S.N.,
Folger,K.R., Kas,A., Lardig,K., Lim,R.M., Smith,K.A., Spencer,D.H.,
Wong,G.K.-S., Wu,Z., Paulsen,I.T., Reizer,J., Saiter,M.H.,
Hancock,R.E.W., Lory,S. and Olson,M.V.
Direct Submission
Submitted (16-MAY-2000) Department of Medicine and Genetics,
University of Washington Genome Center, University of Washington,
Box 352145, Seattle, WA 98195, USA
Location/Qualifiers
1. 10475
/organism="Pseudomonas aeruginosa"
/strain="PA01"
/db_xref="taxon:287"
153..743
/gene="PA5244"
153..743
/gene="PA5244"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08629.1"
/db_xref="GI:9951554"
/translation="MMLOEFLDQFGYFALAGTFEGERTLLVLAGEFLAERGWMQDVT
ILAFLGSYAGDQWYFLGRHGRRLIARFPKQKLGDAKDALDHVRHPLDMLVLSFRFV
YGLRTPVPVAVIGSGYPPARYLLINGALVMAALGSAAYVYGVLEGMGIRKRYE
LWVLGGLIVIGLLIMLRKRKTRGNGAGNADKSE"
908..1576
/gene="PA5245"
908..1576
/gene="PA5245"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08630.1"
/db_xref="GI:9951555"
/translation="MHRKVAIVILSGCGVYDGAIEHRSVITLRLSORGAEOCFAPNI
AOHHVYVHLLTGEEMPRESNRILVESARIRAREVEDLAEAREDDALIVPGFCAAKNL
SDRAINGAOQVQPDVIALAKARAEAKRPGULICIRPAMAKIYKGAVYQVTCGNDADT
ADALKQWGAHIDIPYDEIVEDSARKLVTPAYMALOSTIAEASGINKRLDRLVELVD
QDKA"
1672..2145
/gene="PA5246"
1672..2145
/gene="PA5246"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08631.1"
/db_xref="GI:9951556"
/translation="MSETPSPALPDDVHQLAVDFEORIPFNRMLGTRVASTGRRVTL
DLPEKDELIGNFVQGLIHGVTISLDDVTGGAALIGALRRHHELPHEHMAKLSKIG
TIDRLVDYLRGRGRFTTAHVPLRSKNVAVYSELHSDGTLVAAGTGYLCG"
complement(2130..2612)
/gene="PA5247"
complement(2130..2612)
/gene="PA5247"
/codon_start=1
/transl_table=11
/product="conserved hypothetical protein"
/protein_id="AAG08632.1"
/db_xref="GI:9951557"
/translation="WRWIDADACPRAKELVCFALKRKLEWMMVAGQOVYPPFAC
VRLVAVSGMDAADYVEADPEPDALICSDVPLADLILKQVAAADLPKRFERDARKM
GDKTAMNLMADLDDQGMGGCAQPYGERDROAFANALDRLRLRLQREADLRASQPHR
"

gene
CDS
complement(2679..4580)
/gene="PA5248"
/codon_start=1
/transl_table=11
/product="hypothetical protein"

/protein_id="AA008633.1"
 /db_xref="gi:9951558"
 /translation="MHPRKILMLPLVCAISLGLADPTAVDPSQALHLSTYLAADP
 PVADKQIDPSEYOVEFVGNLQALVLTLPMPREARELERGVALSQALEQRLPDR
 DVALQARNLEAREVADIVYOVPPATTPDPSRAPIYAQOCAICHDAGKGGPGAGILP
 EPPAPALDPRORLDLSTLDLRNVNIGLVAGCTDMPAPADQDEORMDLASYGLSA
 GSAQPKAHAYPLATILATQTPAEVAHEHGEAASEFRRLRHPLEDRPGQILDTYNA
 ATLDKSEAVYACQEDRDOADVLTSAVYLEGFELVSESDYNQADLRSTREKQIMAYRQAL
 LBDGPEYQVAAQGLDLAKGLAEAKKQJCGSLSISFVVALILITLREGVAILLVA
 AILATLRNIGQSAVARGVHGWGLVGFVGEFATWALAAVYLDIGAGRLHEGPFSLTA
 CVMWILGWMDHRRHAAAMQDYIRSSLVGGGGRGFVAVLFEVYRELEVEYLTFFST
 LMVQGPAGNHNHIVIGGAATAVLLILGLMWILLRGSARLPGLFESINALLCALSVYF
 ACHGYALQDEACVICTRPVPEFDFMGLIKADAVLSAQAMALVALILYGRSIVER
 RRAANAAD"
 complement(4685. . 5314)
 /gene="PA5249"
 complement(4685. . 5314)
 /codon_start=1
 /gene="PA5249"
 /codon_start=1
 /product="hypothetical protein"
 /protein_id="AA008634.1"
 /db_xref="gi:9951559"
 /translation="MLVSTWFAFLACMAISLSPGAGIASMSGGLQYGFARGYNAL
 GLQIDGLIQLAVIAGVAGALLATSAISLTKMGEVAVLYVLAIRQWQAPQALSTDG
 EEPRLPLQIPLAVAGFLVANAENKAVIFEMALVDFIDPHOVLQAYLOIMGCTIMVLD
 IYMAVYGTGAARVILVLRSPROOKLVNRTFASLFEVGAAGLATAVRAPL"
 complement(5544. . 6302)
 /gene="PA5250"
 complement(5544. . 6302)
 /gene="PA5250"
 complement(5544. . 6302)
 /codon_start=1
 /gene="PA5250"
 /codon_start=1
 /transl_table=11
 /product="conserved hypothetical protein"
 /protein_id="AA008635.1"
 /db_xref="gi:9951560"
 /translation="MEWLASPEIWAFLFTLTALIVYICIDNIIFISILVGRLPAPQAP
 KAREFGLAMKQTRLILLLSTIWWRLITTDILFHLFGGIGISQRLILFFGGLFLIPST
 MEIYSLGAEAEQKSGKGVGFMGIYQALIIDIFSLDSVTAVGALVONVYFAM
 IYISGVAMVLAAGTISDIIEKHPISLKMALISFLIYVGTAVVAESFEVHVPKGYVFM
 AFSLVAELNIRIRGALARKRGEPEPFLRKSPD"
 complement(6309. . 6887)
 /gene="PA5251"
 complement(6309. . 6887)
 /gene="PA5251"
 complement(6309. . 6887)
 /codon_start=1
 /gene="PA5251"
 /codon_start=1
 /transl_table=11
 /product="hypothetical protein"
 /protein_id="AA008636.1"
 /db_xref="gi:9951561"
 /translation="MEDQLVLTQTSWSEPLVYAMQILILLVAMILQRLVRYTRIGE
 RIVPEPLLEPLRGLRLMWINVASLIVLLEHGVASAVLTATLGFANAVAAFYIW
 SVLSMFCALLFIVTGPRIQRIDREVELIDSADPCVTGRVAVAINLLVTTLEETLDESP
 GPLQIPNSLFFQKAVRWRGRGFGSFTKNGK"
 complement(6887. . 8803)
 /gene="PA5252"
 complement(6887. . 8803)
 /gene="PA5252"
 complement(6887. . 8803)
 /codon_start=1
 /gene="PA5252"
 /codon_start=1
 /transl_table=11
 /product="probable ATP-binding component of ABC
 transporter"
 /protein_id="AA008637.1"
 /db_xref="gi:9951562"
 /translation="WIRLNLITLQRGFORLLAEALITLHAGQAGILGANGAKSSILF
 ALRLQDQDGDCLIPADWRILQMRQEVPTLRLADVLDIDSDRLREIOALAAVE
 AAHGSALARLHETLDNADGYTADARAKRLLAGIGESSEMERARVDEFGSGWMRLNL
 AOALMCPDILLDEPTNHLDDILAITMLEWLGKPGETIVLISHDPDFLDADVVDVHL
 LENKRLTILRGYSAFERTBAERLAAQQAAYEQQAORAHMESFIARFAKAKARQA
 OSRLKLERLEELAPAHVDSPENSFPRESQISRPILLDDEGSLGIDKXAVLEKYLQ
 LYPARICILDPKNGAKSTLIKTLAGDPELGGHLRGENLAIQTPAQHQLDSLDQA
 SPLHLQIATAGEERQTLKDFLGGDFRGVAVDEPVYINFGGKARALALALAMQPN
 LLLDLEPTNHLIDLEPGLTMAIQDFSGAVLVVYSHDRLIKSTTDEFLIADGRVYF

[illegible]

* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.

* 1 296: contig of 296 bp in length
* 297 1049: contig of 753 bp in length
* 1050 1812: contig of 763 bp in length
* 1813 2420: contig of 608 bp in length
* 2421 2953: contig of 533 bp in length
* 2954 3645: contig of 692 bp in length
* 3646 4766: contig of 1121 bp in length
* 4767 5296: contig of 530 bp in length
* 5297 5584: contig of 288 bp in length
* 5585 6454: contig of 870 bp in length
* 6455 6853: contig of 399 bp in length
* 6854 7566: contig of 713 bp in length
* 7567 8253: contig of 687 bp in length
* 8254 9058: contig of 805 bp in length
* 9059 9678: contig of 620 bp in length
* 9679 10272: contig of 594 bp in length
* 10273 10890: contig of 618 bp in length
* 10891 11695: contig of 805 bp in length
* 11696 12677: contig of 982 bp in length
* 12678 13130: contig of 453 bp in length
* 13131 13878: contig of 748 bp in length
* 13879 14026: contig of 148 bp in length
* 14027 14869: contig of 843 bp in length
* 14870 15738: contig of 863 bp in length
* 15739 16426: contig of 688 bp in length
* 16427 17161: contig of 735 bp in length
* 17162 17938: contig of 777 bp in length
* 17939 18715: contig of 777 bp in length
* 18716 19493: contig of 778 bp in length
* 19494 20212: contig of 719 bp in length
* 20213 20780: contig of 568 bp in length
* 20781 21499: contig of 719 bp in length
* 21500 21570: contig of 71 bp in length
* gap of unknown length

* 21571 22378: contig of 808 bp in length
* 22379 22940: contig of 562 bp in length
* 22941 23943: contig of 1003 bp in length
* 23944 24650: contig of 707 bp in length
* 24651 25686: contig of 1036 bp in length
* 25687 25820: contig of 134 bp in length
* 25821 26268: contig of 448 bp in length
* 26269 27152: contig of 884 bp in length
* 27153 28057: contig of 905 bp in length
* 28058 29002: contig of 945 bp in length
* 29003 29991: contig of 989 bp in length
* 29992 30780: contig of 789 bp in length
* 30781 31514: contig of 734 bp in length
* 31515 32152: contig of 638 bp in length
* 32153 32933: contig of 781 bp in length
* 32934 33042: contig of 109 bp in length
* 33043 34310: contig of 1268 bp in length
* 34311 34992: contig of 682 bp in length
* 34993 35821: contig of 829 bp in length
* 35822 37013: contig of 1192 bp in length
* 37014 38147: contig of 1134 bp in length
* 38148 38852: contig of 705 bp in length
* 38853 39840: contig of 988 bp in length
* 39841 40874: contig of 1034 bp in length
* 40875 42387: contig of 1513 bp in length
* 42388 43258: contig of 871 bp in length
* 43259 44365: contig of 1107 bp in length
* 44366 45674: contig of 1309 bp in length
* 45675 47108: contig of 1434 bp in length
* 47109 48115: contig of 1007 bp in length
* 48116 49871: contig of 1756 bp in length
* 49872 51661: contig of 1790 bp in length
* 51662 53728: contig of 2067 bp in length
* 53729 56784: contig of 3056 bp in length
* 56785 59220: contig of 2436 bp in length
* 59221 62274: contig of 3054 bp in length.
* location/Qualifiers
* 1. .62274

FEATURES
source

```
/organism="Mus musculus"
/db_xref="taxon:10090"
/clone="RG-MBAC_40H2"
BASE COUNT      16965 a 13547 c 13251 g 18170 t      341 others
ORIGIN

alignment_scores:
  Quality: 270.50      Length: 131
  Ratio: 3.039         Gaps: 7
  Percent Similarity: 67.939      Percent Identity: 61.069

alignment_block:
US-09-466-935-2 x AC020833  ..

Align seg 1/1 to: AC020833 from: 1 to: 62274

10213 TTGATATGCTTGGGC..... 10227
98 ULYSerLeuAlaSerThrGln..SerArgHisLeuPhe..... 111
10228 .....TGTTCCCCCACCMAATCATCTTGACATTTCCCATGTGTT 10270
112 ...GlnArgAlaValPheValAsnLeuThrAsnProLysSerIleValPh 127
10271 GTTAGCCGTCCTGTG.....ATTCTCAATCCAAAGATATGTGTTC 10311
127 eLeuAlaAlaLeuPheProGlnPheIleMetProGlnInProGlnLeuM 144
10312 TGGGGGGCGCTATT...TGCATATCATCATGCGCCACAGCGGACACGTA 10358
144 eGlnIYrIleValLeuGlyValThrThrIleValIleValAspIleIleVal 160
10359 TGCAGTATATGCTGCTGCGCTCCGCACTAT..GTGTGATATATTATGTG 10407
161 MetIleGlyTYrAlaThrLeuAla..GlnArgIleAlaLeuThrIle...L 176
10408 ATGATCGGTTCAGCCACCTTGCTTCANAGGATGCTCTATGACATTAAAN 10457
176 YsgIyProLysGlnMetLysAlaLeuAsnLysIlePheGlySerLeuPhe 192
10458 AGGACCAACAGCAGATGAGAGCGCTGATATAGATTTCGCGCTCGTGTTC 10507
193 MetLeuValIleGlyAlaLeuLeuAlaSerAlaArgHisAla 205
10508 ATGCTGTGGGAGCGCTGTAGCATCGCGGAGCATGCG 10546

seq_name: gb_baz:AP003001

seq_documentation_block:
LOCUS      AP003001      345783 bp      DNA      BCT      03-FEB-2001
DEFINITION Mesorhizobium loti DNA, complete genome, section 8/21, complete
sequence.
ACCESSION  AP003001 BA0000012
VERSION    AP003001.1 GI:11994976
KEYWORDS   HTG.
SOURCE      Mesorhizobium loti (strain:MAFF303099) DNA.
ORGANISM   Mesorhizobium loti
Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
Phyllobacteriaceae; Mesorhizobium.

REFERENCE
AUTHORS    Kaneo,T., Nakamura,Y., Sato,S., Asamizu,E., Kato,T., Sasamoto,S.,
            Watanabe,A., Idegawa,K., Ishikawa,A., Kawashima,K., Kimura,T.,
            Kishida,Y., Kiyokawa,C., Kohara,M., Matsumoto,M., Matsuno,A.,
            Mochizuki,Y., Nakayama,S., Nakazaki,N., Shimpo,S., Sugimoto,M.,
            Takeuchi,C., Yamada,M. and Tabata,S.
TITLE       Complete genome structure of the nitrogen-fixing symbiotic
            bacterium Mesorhizobium loti
JOURNAL     DNA Res. 7, 331-338 (2000)
REFERENCE   2 (bases 1 to 345783)
AUTHORS     Kaneo,T.
```

```
TITLE      Direct Submission
JOURNAL     Submitted (05-DEC-2000) Takakazu Kaneko, Kazusa DNA Research
            Institute, The First Laboratory for Plant Gene Research; Yana
            1532-3, Kisarazu, Chiba 292-0812, Japan
            (E-mail:kaneko@kazusa.or.jp)
            URL:http://www.kazusa.or.jp/rhizobase/, Tel.81-438-52-3935,
            Fax:81-438-52-3934)
FEATURES
source      1..345783
            /organism="Mesorhizobium loti"
            /strain="MAFF303099"
            /db_xref="taxon:381"
BASE COUNT  64252 a 111029 c 106519 g 63983 t
ORIGIN

alignment_scores:
  Quality: 255.50      Length: 205
  Ratio: 1.907         Gaps: 4
  Percent Similarity: 65.366      Percent Identity: 33.171

alignment_block:
US-09-466-935-2 x AP003001  ..

Align seg 1/1 to: AP003001 from: 1 to: 345783

1 MetThrLeuGluTYrTYrPheAlaTYrLeuLeuThrSerIleIleLeuTh 17
137079 ATGTGCGTGGAGACTCTATGCCGCTTATGTTCTCCGTCGATCGATCAT 137128
17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnH 34
137129 CCTGTGCGCGGGAGCGAGCGTCACACATGATCATCGCCAAAGCATCCGCC 137178
34 tsgIyTYrProAlaGlyIleGlyValTYrCystIrpAlaSerAspArgThgIy 50
137179 ACGGAACCCGCGGC...GGCCTCGCCAATGTCGCCGCGGACGCGCGGG 137225
51 AspSerTYrCysAlaGlyTYrPArgGlyValGly.....ThrLeuPhe 65
137226 CTTCGCATCATGATGCCATTTGTGGCATGCGCTCCAAACGCTGATCCG 137275
65 rArgSerValIleAlaPheGluValLeuLysTYrPalAglValAlaTYrL 82
137276 GGGCATGGGCGCAGCTGTTGAATGGTCAGCGTATGCGCGCCGCTATTC 137325
82 euIleTYrLeuGlyIleGlnInTYrPArgAlaIleAlaIleAspLeu 98
137326 TGATCTGGATGGGCGTGCAGATGTTCCGCTCCAAGGCGACGCTGAATGCC 137375
99 LysSerLeuAlaSerThrGlnSerArgHisLeuPheGlnArgAlaVal 115
137376 GACGGAAACCGCT...AGGAACCGCGCGCGCTTCTCCGACAGGCGCT 137422
115 lPheValAsnLeuThrAsnProLysSerIleValIlePheLeuAlaLeuP 132
137423 GCTGTGGCGCTGAGCAATCCCAAGACCTGTGTTCTTCGCGCGCTTCT 137472
132 heProGlnPheIleMetProGlnInProGlnLeuMetGlnTYrIleVal 148
137473 TCCCGCACTTATCGCGCGGAGCAACACTATTCGCTCAAAATCGTGTCT 137522
149 LeuGlyValThrThrIleValValAspIleIleValMetIleGlyTYrAl 165
137523 ATGGGCTTCAGCCGCGCATGATCTTGCCCGCCCATGTCGATTCGACTACGC 137572
165 aThrLeuAlaGlnArgIleAlaLeuTYrIleLysGlyProLysGlnMetL 182
137573 GCTGCGCGCGCGCGCGCGGACGCGCTCTGTGCGGC...AGCCGCAATCA 137619
182 ysaAlaLeuAsnLysIlePheGlySerLeuPheMetValGlyAlaLeu 198
137620 AGCTCATGTCCAGGATCAAGCGGACGCTTGTGTGGTGGCGGCGGCTGTGG 137669
```

199 LeuAlaSerAlaArg 203
|||||
:::|
137670 CTGGCGTTTCGAGG 137684

1
2
3
4

OM of: US-09-466-935-2 to: N_Geneseq_0401.* out_format : pfs

Date: May 6, 2001 3:43 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

```
-MODE=framed-p2n.model -DPV=xlp  
-O=/cgm2_1/USPTO.spool/US09466935/runatc_03052001.075710.14843/app_query.fasta.1.529  
-DB=N_Geneseq_0401 -QFMT=fastap -SUFFIX=ring -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -XGAPOP=10.000 -XGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blissum62  
-TRANS=human40.ccl -LIST=45 -DOCALLIGN=200 -THR_SCORE=pcr  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pis  
-NORM=ext -MINLEN=0 -MAXLEN=2000000000  
-USER=US09466935_@CGN1_1_213 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPHY  
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-466-935-2  
Query length: 205  
Database: N_Geneseq_0401.*  
Database sequences: 678276  
Database length: 291890651  
Search time (sec): 172.640000
```

```
Score List:
Sequence      Strd Orig      ZScore      EScore      Len      Documentation
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:294405 + 1061.00 2192.23 3.6e-114 1200 I E
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48442 + 1061.00 2191.94 3.7e-114 1231 I E
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52689 + 195.50 389.17 9.7e-14 639 I E
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:R06767 + 184.00 363.80 2.5e-12 720 I E
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V08533 + 184.00 363.80 2.5e-12 720 I S
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C74615 + 144.00 286.62 5.0e-08 401 I H
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52688 + 126.50 244.32 1.1e-05 672 I E
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52690 + 126.50 244.75 1.1e-05 588 I E
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A61501 + 123.50 227.99 9.3e-05 1668 I A
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A61502 + 123.50 227.99 9.3e-05 1668 I A
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48443 + 122.00 232.44 5.2e-05 840 I E
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52691 + 113.00 216.68 0.0004 636 I E
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:R96861 + 89.50 152.99 1.39 2374 I T
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V27579 + 85.50 145.01 3.87 2289 I H
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:X13048 + 85.00 132.27 19.84 6590 I F
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A81730 + 83.50 125.53 47.12 9133 I N
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:V52687 + 82.50 143.86 4.49 1441 I H
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V27578 + 82.00 142.46 5.37 1635 I H
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V24288 + 82.00 137.81 9.75 2265 I H
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:Z44289 + 82.00 137.81 9.75 2265 I H
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A51307 + 82.00 109.73 357.43 28690 I
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:X13075 + 80.50 132.45 19.38 2769 I N
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:P44697 + 80.50 117.24 136.35 10953 I
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V74373 + 80.00 138.55 8.87 1452 I A
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:X19826 + 80.00 137.82 9.74 1551 I A
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:X12151 + 79.50 147.82 5.34 924 I N
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:R081959 + 79.50 137.62 10.00 1437 I F
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:R21233 + 79.50 131.52 21.85 2494 I F
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A35111 + 79.50 131.52 21.85 2494 I F
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:R21235 + 79.50 111.97 268.23 14607 I
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A35113 + 79.50 111.97 268.23 14607 I
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V05287 + 79.50 98.49 1.5e+03 49377 I C
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:X56244 + 79.50 123.07 20.35 2158 I C
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:R12223 + 78.50 134.75 14.45 1542 I N
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A95944 + 78.50 112.22 259.75 11820 I
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V27577 + 78.00 129.34 28.91 2288 I H
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:R80948 + 78.00 117.93 34.62 2598 I H
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V52731 + 78.00 133.00 235.07 10023 I
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:X98132 + 77.50 144.83 3.96 513 I NU
```

```
/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:Q43846 + 77.50 111.22 295.08 10704  
/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:Q43848 + 77.50 111.08 300.57 10844  
/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:Q43844 + 77.50 111.40 327.86 11529  
/SIDS2/gcgdata/geneseq/geneseq/NA1993.DAT:Q43847 + 77.50 109.84 352.45 12132  
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V52660 + 77.00 138.27 9.19 844 I
```

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:Z94405

seq_documentation_block:

ID Z94405 standard; DNA; 1200 BP.

XX Z94405;

XX 18-JUL-2000 (first entry)

XX E. coli rhtB gene imparting homoserine resistance.

XX Homoserine resistance; rhtB gene; L-homoserine; L-alanine;

KW L-Isoleucine; L-valine; L-threonine; ss.

XX Escherichia coli.

XX Key Location/Qualifiers

FT CDS 557..1171

FT /*tag= "a
/note= "a DNA sequence corresponding to the coding
region (minus the stop codon) is
specifically claimed in Claim 3"

EP994190-A2.

PD 19-Apr-2000.

XX 20-SEP-1999; 99EP-0118581.

XX 13-OCT-1998; 98RU-0118425.

(AJIN) AJINOMOTO CO INC.

XX Llysylts VA, Zakataeva NP, Aleoshin VV, Belareova AV;

XX Tokhmakova II;

XX WPI; 2000-273530/24.

XX DR P-PSDB; Y79298.

XX Novel RhtB protein, useful for generation of L-homoserine resistance in
Escherichia bacteria and large-scale production of e.g. L-homoserine
and L-alanine

XX Claim 3; Page 10-11; 14pp; English.

This is the DNA sequence the novel rhtB gene of Escherichia coli
K-12. The rhtB gene maps at 86 min on the E. coli chromosome. The
gene participates in resistance to homoserine. Amplification of
the gene results in an improvement of the amino acid productivity
of E. coli. The invention provides: the RhtB protein (see W79298);
DNA encoding the RhtB, especially nucleotides 557-1171 of the
present sequence; a bacterium, especially of the genus Escherichia,
in which L-homoserine resistance is enhanced by amplifying the
copy number or increasing the expression rate of the rhtB DNA, the
DNA being carried on a multicopy vector or on a transposon; and a
method for producing an amino acid by cultivating the bacterium in
a culture medium to produce and accumulate the amino acid in the
medium, from which it is recovered. The method is used for the
production of L-homoserine, L-alanine, L-isoleucine, L-valine or
L-threonine (all claimed).

Sequence 1200 BP; 285 A; 309 C; 305 G; 301 T; 0 other;

alignment_scores:

Quality: 1061.00 Length: 205
Ratio: 5.176 Gaps: 0


```
84 rpleuaglylleinglntrpargaiaaiaaglyalaalleaspleuylsSer 100
|||||
807 GGCTGGGAATCCAGACAGTGGCGCGCTGGCATTTGACCTTAATTCG 856
101 leuAlaSerThrGlnSerArgArgHisLeupheglnArGaIaValaIphaVa 117
|||||
857 CTGGCTTACTCAATCGCGTGCATATTGTTCACGCCGCGAGTTTGT 906
117 lAsnLeuThrAsnProLysSerIleValaPheLeuAlaIaLeupheProc 134
|||||
907 GAATCTCACCAATCCAAAAGATTTGTCTCGCGCGCTATTTCGCG 956
134 lPheIleMetProGlnGlnProGlnLeuMetGlnTyrIleValaIleuGly 150
|||||
957 AATTCAATCATGCCGACAGCGCCGCAACTGATGATATATGCTCTGGCC 1006
151 ValThrThrIleValaValaAspIleIleValaMetIleGlyTyrAlaThrLe 167
|||||
1007 GTACCCACTATTGTGGTGGATATTATGTGATGATCGGTTACGCCACCT 1056
167 uAlaGlnArgIleAlaLeuThrIleLysGlyProLysGlnMetLysAlaL 184
|||||
1057 TCCTCAACGAGATTGCTATGATGATTAAAGCAACCAAGCATGAAGCGCG 1106
184 euAsnLysIlePheGlySerLeuPheMetLeuValGlyAlaLeuLeuAla 200
|||||
1107 TGAATAGATTTCGCTCGCTGTTGTTATGCTGGTGGAGCGCTGTAGCA 1156
201 SerAlaArgHisAla 205
|||||
1157 TCGCGAGGACATGCG 1171

seq_name: /SIDS2/gcdata/geneseq/geneseqn/NA2000.DAT.A52689
seq_documentation_block:
ID A52689 standard; DNA; 639 BP.
XX
XX A52689:
XX
XX 03-JAN-2001 (first entry)
XX
XX Escherichia coli yeast gene.
XX
XX DE Escherichia coli yeast gene.
XX
XX E. coli: yeast gene; amino acid production; excretion protein gene;
XX amino acid excretion protein; ds.
XX
XX OS Escherichia coli.
XX
XX FH Key Location/Qualifiers
XX FT CDS 1..639
XX FT /*tag= a
XX FT /product= "Yeast"
XX
XX PN EP1016710-A2.
XX
XX PD 05-JUL-2000.
XX
XX PF 17-DEC-1999; 99EP-0125263.
XX
XX PR 30-DEC-1998; 98RU-0124016.
XX PR 09-MAR-1999; 99RU-0104431.
XX
XX PA (AJIN ) AJINOMOTO CO INC.
XX
XX PI Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
XX Tokimukova II;
XX
XX DR WPI: 2000-414802/36.
XX
XX P-PSDB: B01787.
XX
XX Increased production of L-amino acids by an Escherichia bacterium
XX PT comprises increasing the expression amount of an L-amino acid excretion
XX PT protein -
```

```
XX
PS Disclosure: Page 19-20; 29pp; English.
XX
XX The present sequence is the yeast gene (an excretion protein gene) of
XX Escherichia coli. The amino acid excretion protein produced from this
XX gene is involved in the production of amino acids, and an increase in its
XX expression leads to an increased accumulation of amino acids in the cell.
XX In this case, an increase in lysine, alanine, valine, histidine,
XX CC isoleucine, glutamic acid and proline is achieved if multiple copies of
XX the gene are transfected into a bacterium. The bacterium used is E. coli.
XX
XX Sequence 639 BP: 142 A; 128 C; 161 G; 208 T; 0 other;
XX
XX alignment_scores:
XX Quality: 195.50 Length: 213
XX Ratio: 1.387 Gaps: 10
XX Percent Similarity: 66.197 Percent Identity: 28.638
XX
XX alignment_block:
XX US-09-466-935-2 x A52689 ..
XX
XX Align seg 1/1 to: A52689 from: 1 to: 639
XX
XX 3 leuGlnTPrThrPheAlaTyrLeuLeuThrSerIleIleLeuThrLeuSe 19
XX |||||
XX 22 CTGATTTACTGG...ACCTATCTGTTGGGCGCATTTTATGTTGTGTGT 68
XX |||||
XX 19 rProGlySerGlyAlaIle.....AspThrMetThrThrSerL 32
XX |||||
XX 69 GCCAGGCGCCAAATACCTGTTTGTACTCAAAATATAGGAGTACGAGCGTA 118
XX |||||
XX 32 euAsnHisGlyTyrProAlaGly..GlyValTyrCysTyrPAlaSerAsp 47
XX :::: |||||
XX 119 TGAAGCGCGTTATCTCTCGCGCTGCGGTGATTT..... 153
XX |||||
XX 48 ArgThrGlyAspSerTyrCys.....AlaGlyTyrArgGlyValGlyTh 62
XX |||||
XX 154 ..ATTGGGAGTGGGATGATGATGTTCTGGCATGGGCTGAGTGGCGAC 200
XX |||||
XX 62 lLeuPheSerArgSerValIleAlaIlePheGluValaLeuLysTyrPAlaGly 79
XX |||||
XX 201 ATTATTTAAAGACACACCCGATATATTCACATGTCAGTTATCTGTGTC 250
XX |||||
XX 79 lAlaIleTyrLeuIleThrPleuGlyIleGlnGlnTrpAlaIaIaGlyAla 95
XX |||||
XX 251 CGTTTATTTGCTCTATCTGGGAGTAAATTCTTTCAGCGACC..... 294
XX |||||
XX 96 lIleAspLeuLys.....SerLeuAlaSerThrGlnSerArgArgHl 109
XX |||||
XX 295 .....CTGAAGGGTAAATATATGAGGCGCAATCCGATGAGCCCCATA 338
XX |||||
XX 109 s.....LeupheGlnArgAlaValaIlePheValaAsnLeuThrAsnProLys 124
XX :::: |||||
XX 339 CGGTCCTATTTTAAACGCGCTTAATTTGAGCTCATGCTAATCCGAAG 368
XX |||||
XX 124 erIleValaPheLeuAlaIaLeupheProcGlnIleMetProGlnGln 140
XX :::: |||||
XX 389 CCATTTTGTCTATGTCGCTTTTTCGTACAGTTATGATGATTAAGCC 438
XX |||||
XX 141 ProGlnLeu...MetGlnTyrIleValaLeuGlyValaIleThrIleValaIva 156
XX |||||
XX 439 CCACATACGGGAATTTCACTTCTATTCTGGCGGAGCGTGAACCTGGT 488
XX |||||
XX 156 lAspIleIleValaMetIleGlyTyrAlaThrLeuAlaGlnArgIleAlaL 173
XX |||||
XX 489 GAGTTTCTGATTTTGACCTCTCGATATATATCTGCTGCTTTGTCACGC 538
XX |||||
XX 173 euThrIleLysGlyProLysGlnMet...LysAlaLeuAsnLysIlePhe 188
XX :::: |||||
XX 539 AGTACATACGTACCAAAAGAACTGGCTAAAGTTGGCACTCACTGATTT 568
XX |||||
XX 189 GlySerLeuPheMetLeuValaIleAlaLeuLeuAlaSer 201
```

```

seq.name: /SIDS2/gcgdata/geneseq/geneseqn/A1196.DAT:T06767
seq-documentation_block:
ID T06767 standard; DNA: 720 BP.
XX
AC T06767;
XX
DT 01-JUL-1996 (first entry)
XX
DE Mel-linked mlga gene.
XX
KW Marine mela; selectable marker; oyster larva settlement;
KW pigmentation; melanin; p-hydroxyphenylpyruvate-hydroxylase;
KM mlga gene; ss.
XX
OS Shewanella colwelliana strain LSTDF.
XX
FH Key Location/Qualifiers
FT RBS 154..159
FT RBS /*tag= a
XX
PN US5474933-A.
XX
PD 12-DEC-1995.
XX
PE 21-MAR-1990; 90US-0496804.
XX
PR 08-NOV-1993; 93US-0148945.
PR 21-MAR-1990; 90US-0496804.
PR 10-NOV-1992; 92US-0374837.
XX
PA (UYMA-) UNIV MARYLAND BALTIMORE.
XX
PI Fuqua WC, Welner RM;
DR WPI: 1996-039515/04.
DR P-PSDB; R87527.
XX
PT Novel gene encoding marine mela from Shewanella - useful as
PT selectable marker in genetic engineering and for inducing larval
PT oyster settlement
XX
PS Example 7; Fig 14; 47pp; English.
XX
CC The 5' end of an open reading frame (T06767) was identified
CC directly downstream of the Shewanella colwelliana mela gene
CC (T06766), and was designated mlga (mel-linked gene). The
CC role of the encoded protein (R87527) was unclear as deletion
CC subcloning in E. coli demonstrated that only mela was required
CC for melanogenesis.
XX
SQ Sequence 720 BP; 180 A; 165 C; 165 G; 210 T; 0 other;
XX

alignment_scores:
Quality: 184..00 Length: 210
Ratio: 1.546 Gaps: 4
Percent Similarity: 56.667 Percent Identity: 27.619

alignment_block:
US-09-466-935-2 x T06767 ..

Align seg 1/1 to: T06767 from: 1 to: 720

9 TyrLeuThrSerIleLeuThrLeuSerProGlySerGlyAlaIle 25
:::||||| ::::::::::::::::::::|||
122 TTTTACGCAGCCTTTTTCGTCAATTACACAGCATGTGTATGAC 171
:::||||| ::::::::::::::::::::|||
25 eAsnThreThrThrSerLeuAsnHisGlyTyrProAlaGlyGlyValr 42
:::||||| ::::::::::::::::::::|||

```

172	ACTGGCCATGACTCTCGTATGAGTATCGGTG.....	205
42	YRCYSTPpalaseraSPaRGTGlyAspserTyrCyslaeGlyTPrp	58
206CGCCGAAC.....TTATGATG	223
59	GlYalGlyThrleuPheSerArgSerValleala.....	70
224	AWGTTGTGAGCTAGCAGGCGTTGCCCTCGTGGCGATGTCGCCGTAAT	273
71PheGluValleuL	75
274	GGGTGTGCCAGTATGCTGTGACTATCCACAACCTCTCGATATTAA	323
75	YSTRPalaglyAlaIaIaTyrleuIleTPrleuGlylleGlnGlnTPrp	91
324	AATGGTGTGGTGGCTCTACTGTGTTACATCGGCATWAGCATGTGGCGG	373
92	AlaIaGlyAlaIle...AspIeuYssrleuAlaIaSerThGlnSerAr	107
374	GCCAAAGGGAATGGCCAACTTGACAAATACCTCCAGTCAGATCAGTAA	423
107	gARgHisleuPheGlnArGAlaValPheValAsnleuThraSProlYss	124
424	TCGAGCCCTAATACCTCAAGCGCTTTGTCAACCCCAATTGCTATCCAAAG	473
124	erlleValPheleuAlaIaIaIeuePheProGlnPheIleuPheProGln	140
474	GTTGGGCTTTATGATCTCGTCGTCGCCCTTTTATCAGCGTTACCA	523
141	ProGlnleuMetGlnTyrIleValleuGlyValThrThrIleValleValas	157
524	GCGATTGCACCACCAATTAATGATTACTGTCAATTATATGATGACAGA	573
157	PleIleValMetIleGlyTyrAlaThrIleuAlaGlnArGileAlaIeUT	174
574	GTTCTCAGCATGCTGCTCTTATGCGACCG.CGGAACACCTTAACACTGT	622
174	rPlleYsglyProLYsglnMetIYsAlaIeuaSnLYsIlePheGlySer	190
623	TTTTAAGTCGAGCGGATACATCAAGTGAATGACCGCATACAGAGGAGT	672
191	leuPheMetIeUValGlyAlaIeuleuAla	200
673	TTAATGATCTGTGTGGCTTATGTTGGG	702
seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:v08533		
seq_documentation_block:		
ID	V08533	standard: DNA: 720 bp.
AC	V08533;	
XX		
DT	11-FEB-1999	(first entry)
XX		
DE	S. colwelliana M1ga coding sequence.	
XX		
KW	M1GA; marine Bacterium; melanin synthesis; marine exopolysaccharide;	
RV	UV blocker; sunscreen; Mela; ss.	
XX		
OS	Shewanella colwelliana.	
XX		
Key		Location/Qualifiers
FT	CDS	167..628
XX		/tag= a
FN	US5846531-A.	
PD	08-DEC-1998.	
XX		
EF	07-JUN-1995.	95US-0476254.
XX		
PR	07-JUN-1995.	95US-0476254.

```
PR 21-MAR-1990; 900S-0496804.
PR 10-NOV-1992; 92US-0974837.
PR 08-NOV-1993; 930S-0148945.
XX
PA (UYMA-) UNIV MARYLAND BALTIMORE.
PI
XX Fugua WC, Weiner RM;
XX
DR WPI; 1999-058995/05.
DR P-PSDB; W73358.
XX
PT Mela melanin protein from marine bacteria - useful as UV blocker in
PS e.g. cosmetics
XX
XX Example 7; Fig 13; 57pp; English.
XX
XX This sequence encodes the Shewanella colwelliana M1ga protein, which is
CC related to the Mela protein of the invention. The invention relates to
CC Mela proteins from marine bacterium of the genus Shewanella, Vibrio or
CC Hyphomonas. The Mela protein is active in catalysing production of
CC melanin synthesis, as well as components of marine exopolysaccharides.
CC The Mela protein can be used as a UV blocker in sunscreen, pigment and
CC dye compositions useful in cosmetic applications, as tints for glass or
CC in paints and coatings for plastics, synthetic resins and fabrics, rubber
CC and wood.
XX
XX Sequence 720 BP; 180 A; 165 C; 165 G; 210 T; 0 other;
SQ

alignment_scores:
    Quality: 184.00      Length: 210
    Ratio: 1.546         Gaps: 4
    Percent Similarity: 56.667   Percent Identity: 27.619

alignment_block:
US-09-466-935-2 x V08533 ..

Align seg 1/1 to: V08533 from: 1 to: 720

9 TyrLeuLeuThrSerIleIleLeuThrLeuSerProGlySerGlyAlaIle 25
   ::||| ||| ::::::::::::::::::::|||
122 TTTTACCGACCTTTTTCGTCTCAATTACACGAGTATGTGTATGAC 171
   ::||| ||| ::::::::::::::::::::|||
25 eAsrThmEthrThrSerLeuAsnHISGlyTyrProAlaGlyGlyValr 42
   ::||| ||| ::::::::::::::::::::|||
172 ACTGGCATGACTCTCGTATGAGTATCGGATG..... 205
42 yrcYsrPalAserAspArgThrGlyAspSerTyrCysAlaGlyTyrParg 58
   ||||| |||
206 .....CGCCGAAC.....TTATGATG 223
59 GlyValGlyThrLeuPheSerArgSerValIleAla..... 70
   ||||| ||| ::::::::::::::::::::|||
224 ATGGTTGGTGTAGTACGAGCGGTGCGCTCGTGGCAGTTCCGCCGTAAT 273
71 .....PheGluValLeu 75
274 GGGTGTCCGAGTATGCTGTAACATCAACAACCTCTGATATTTTAA 323
75 ysrTPAlaGlyAlaIleTyrLeuIleTrrPleuGlyIleGlnGlnTrrParg 91
   ||||| ::||| ::::::::::::::::::::|||
324 AATGGGTGCGTGGGCTCTATCTTGTACATCGCATTTAGCATGTGCGG 373
92 AlaIleGlyAlaIle...AspLeuLysSerLeuAlaSerThrGlnSerAr 107
   ||| ||| ::||| ::||| |||
374 GCCAAGAGGAAATGGCCAACTTGACAAATACCTCTCAGTACGATCACTAA 423
107 gArGHisLeuPheGlnArgAlaValPheValAsnLeuThrAsnProLys 124
   ::||| ||| ::||| ::||| ::|||
424 TCGACCGCTAATTAACCTCAAGCTTGTACCGCAATTGCTAATCCAAAG 473
124 erIleValPheLeuAlaIleLeuPheProGlnPheIleMetProGlnGln 140
   :: ::||| ::||| ::||| |||||
```

```
474 GGTGGGCTTTATGATCTGCTGCTCCGCCCTTTATCAGCGTTGACCA 523
141 ProGlnLeuMetGlnTrrIleValLeuGlyValrThrIleValAla 157
   ||| ::||| |||
524 GCGATTGGACCAACAATTAATGTATGCTCAATTAATATGATGACAGA 573
157 rIleIleValMetIleGlyTyrAlaThrLeuAlaGlnArGleAlaLeu 174
   ::||| ::||| ::||| ::||| ::|||
574 GTTCTTCAGCATGCTTGTCTTATGCGAGCGG.CGGAACCCCTTAACGT 622
174 rPileLysGlyProLysGlnMetLysAlaLeuAsnLysIlePheGlySer 190
   ::||| ::||| ::||| ::||| ::|||
623 TTTTAAGTCGAGCGCATTAACATCAAGTGTGATGACCGCATGACGAGT 672
191 LeuPheMetLeuValGlyAlaLeuLeuAla 200
   ||||| ::||| |||||
673 TTAATGATCTGTGCTTATGCTTATGCTGCG 702

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT.C74616
seq_documentation_block:
ID C74616 standard; cDNA: 401 BP.
XX
AC C74616;
XX
DT 08-FEB-2001 (first entry)
XX
XX Human ORFX ORF171 polynucleotide sequence SEQ ID NO:341.
XX
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
XX vulnery; antipsoriatic; antiparkinsonian; nootropic; neuroprotective;
XX anticonvulsant; osteopathic; antiarthritic; immunosuppressant; cardiant;
XX immunostimulant; thrombolytic; coagulant; vasotropic; antidiabetic;
XX hypotensive; dermatological; immunosuppressive; antiinflammatory;
XX antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
XX antianaemic; gene therapy; cancer; proliferative disorder; hypertension;
XX neurodegenerative disorder; osteoarthritis; graft vs host disease;
XX cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
XX cholesterol ester storage; systemic lupus erythematosus; infection;
XX severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
XX allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
XX bone damage; catilage damage; antiinflammatory disease; coagulation;
XX thrombosis; contraceptive; ss.
XX
XX Homo sapiens.
XX
XX WO200058473-A2.
XX
XX 05-OCT-2000.
XX
XX 31-MAR-2000; 2000MO-US08621.
XX
XX 31-MAR-1999; 99US-0127607.
XX 02-APR-1999; 99US-0127636.
XX 05-APR-1999; 99US-0127728.
XX 30-MAR-2000; 2000US-0540763.
XX
XX (CURA-) CURAGEN CORP.
XX
XX Shinkets RA, Leach M;
XX
XX WPI; 2000-602362/57.
XX
XX P-PSDB; B40407.
XX
XX Novel nucleic acids and peptides derived from open reading frame X,
XX useful for treating e.g. cancers, proliferative disorders,
XX neurodegenerative disorders and cardiovascular disease -
XX
XX Claim 5; Page 607-608; 5507pp; English.
XX
XX C74446 to C77606 encode the proteins given in B40237 to B43397, which
XX represent the human ORFX open reading frames 1 to 3161. The ORFX
XX sequences have activities such as: cytostatic; hepatotropic; vulnery;
```

CC antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
 CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
 CC cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;
 CC dermatological; immunosuppressive; antineoplastic; antibacterial;
 CC antiviral; antifungal; antineumatic; antithyroid; antidiabetic. The
 CC sequences can be used for determining the presence of or predisposition
 CC to, or preventing or treating pathological conditions associated with an
 CC ORF-associated disorder. The nucleic acids can be used to express ORF
 CC proteins in gene therapy vectors. The proteins and nucleic acids may be
 CC used to treat cancers, proliferative disorders, neurodegenerative
 CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
 CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
 CC storage, systemic lupus erythematosus, severe combined immunodeficiency
 CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
 CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
 CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
 CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.

XX
 KW E. coli; yahn gene; amino acid production; excretion protein gene;
 KW amino acid excretion protein; ds.
 XX
 OS Escherichia coli.
 XX
 FH Key Location/Qualifiers
 FT CDS 1..672
 FT /tag= a
 FT /product= "yahn"

alignment_scores:

Quality: 144.00 Length: 129
 Ratio: 1.694 Gaps: 4
 Percent Similarity: 65.891 Percent Identity: 31.008

alignment_block:

US-09-466-935-2 x C74616 ..

Align seg 1/1 to: C74616 from: 1 to: 401

79 ALaAlaIreuleuIetrpleugliIlegInItrParAlaIaIaI 95
 |||||
 7 GCGGCCCTACCTGCTGTAACCTGGCCCTATGCCACTGGCGCTCGC 56
 |||||
 95 alleAspleuLysSerleuAlaSerThrgInSerArgArgHISleuPhag 112
 |||||
 57 CTTTGCAATGACAGACACGCCGACAGTGGACCGCGCGACGCTATCC 106
 |||||
 112 InrgrAlaValpHeValAsnleuThraSProlYsSerleValpHeleu 128
 |||||
 107 TGGGTGGCTTCTTGCAATCTTAACCCCAAGCTGACATTTCTTC 156
 |||||
 129 ALaAlaLeupheProGInpHeIleMetPro.....GInGInProGInle 143
 |||||
 157 CTGGCCCTCTCGCTCATTCCTTAACGCCAGCGGCGACCGCCGCCCTT 206
 |||||
 143 uMetGInTrIleValleuGlyValThrThrIleValAlaSPIleIleY 160
 |||||
 207 G...CAGATGCTGTACTGAGCGCGCTGTTCAMGCGCATGACGCTGCAG 253
 |||||
 160 alMetIleGlyTrAlaThrleuAlaGIn....ArgIleAlaLeuTrp 174
 |||||
 254 TGTGTGGCTGTATGGCTGTGGCAATGTGTTCGTGTGCAGAG... 300
 |||||
 175 IleYsGlyProLysGInMetLysAlaLeuAsnLysIlePhagLysSerle 191
 |||||
 301 GTGAGTCGCGACAGTCGTGCAAACTGCTGCGACGAGCTTTGCCACGCG 350
 |||||
 191 uPHeMetLeuValGlyAlaLeuLeuAlaSerAlaArg 203
 |||||
 351 CTTTGCGGCGCTGGGCTTGACCTTGCGCTTTCGCGCAG 387
 |||||

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52688

seq_documentation_block:

ID A52688 standard; DNA; 672 BP.

AC A52688;

DT 03-JAN-2001 (first entry)

DE Escherichia coli yahn gene.

Sequence 672 BP; 130 A; 154 C; 178 G; 210 T; 0 other;

alignment_scores:

Quality: 126.50 Length: 201
 Ratio: 1.037 Gaps: 6
 Percent Similarity: 60.697 Percent Identity: 21.891

alignment_block:

US-09-466-935-2 x A52688 ..

Align seg 1/1 to: A52688 from: 1 to: 672

1 MethrIleuGInuTrpPheAlaTrIleuLeuThr.....SerI 14
 ::|||
 34 ATCACTATGATCCTTTGCATGCCGTTTACCTGACCGTAGACTGTTGCT 83
 |||||
 14 eIleLeuThrIleuSerProGlySerGlyAlaIleAsnTrMetThrThrs 31
 |||||
 84 GATTACTTTTATTAACCGGAGGCAATCTCTTGTGTGATGCAACCA 133
 |||||
 31 eIleuAsnHISGlyTrProAlaGlyGlyValTrYcStrPalasAsp 47
 |||||
 134 GCCGTGCTTCGCGTGCAGCGCA...GGGTGCTGACCGGCGTGGCGTG 180
 |||||
 48 ArgThrGlyAspSerTrYcysAlaGlyTrParG.....GlyValGlyTh 62
 |||||
 181 GCGCTGGGCAATGATTTTATTCGGGTTGGGTTGTGTGCTTGCAAC 230
 |||||
 62 rIleuPheSerArgSerValIleAlaPheGlyuValleuLysTrPalasGly 79
 |||||

```
221 GCTAATACGCACTGTGAGAGATTTTTCGCTTATCAGAAATCGCGCG 280
79 1aAlaTyrLeuLeuTyrPheLeuGlyIleGlnGlnTrrParGAlaIaGlyAla 95
281 GCGCTTATCTCTTATGTTGGCTGGTCCAGCATGCGCGCCAGCACA 330
96 1LeaSplLeuYsSerLeuAlaSerThrGlnSerArgArgHis.....Le 110
331 CCGCAAAATGACACACTACACAACACGATTAAGCGCCCTGGTATGCTT 380
110 uPheGlnArgAlaValPheValAsnLeuThrAsnProLysSerIleValP 127
381 TTTTCGCGCGGATTAATACGATCTCTACCCGCAACCGCATTTTAT 430
127 heLeuAlaIaLeuPheProGlnPheIleMetProGlnInPro..... 141
431 TTTTATCAGATTTTCTCAGTACATTAATGCGCAACACCAACATGG 480
142 ..GlnLeuMetGlnTyrIleValLeuGlyValThrThrIleValAla 157
481 GCACGTTAATGCGCTGGCGGGGATTTGCTCGCATCAATATCTGGCG 530
157 PileIleValMetIleGlyTyrAlaThrLeuAlaGlnArgIleAlaLeu 174
531 AGTTTTCCTTAGTCAGCGGTTTCTTTGCCCGCTGCTGCTGCTTAT 579
174 rPileLysGlyProLysGlnMetLysAlaLeuAsnLysIlePheGlySer 190
580 .....GGGCGTATGCAACGCGTTCGCCAGTGGGTATGTGTGCA 618
191 Leu 191
619 AT 621

seq_name: /SID52/gcgdata/geneseq/geneseqn/NA2000.DAT:A52690
seq_documentation_block:
ID A52690 standard; DNA; 588 BP.
AC A52690;
DT 03-JAN-2001 (first entry)
XX
DE Escherichia coli yf1k gene.
XX
KW E. coli; yf1k gene; amino acid production; excretion protein gene;
amino acid excretion protein; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..588
FT /tag= a
FT /product= "yf1k"
XX
PN EPI016710-A2.
XX
PD 05-JUL-2000.
XX
PF 17-DEC-1999; 99EP-0125263.
XX
PR 30-DEC-1998; 98RU-0124016.
PR 09-MAR-1999; 99RU-0104431.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
PI Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
Tokmakova IL;
XX
DR WPI: 2000-414802/36.
DR P-PSDB: B01788.
XX
XX Increased production of L-amino acids by an Escherichia bacterium
```

```
PT comprises increasing the expression amount of an L-amino acid excretion
PT protein -
XX
XX Disclosure; Page 22; 29pp; English.
PS
XX
XX The present sequence is the yf1k gene (an excretion protein gene) of
CC Escherichia coli. The amino acid excretion protein produced from this
CC gene is involved in the production of amino acids, and an increase in its
CC expression leads to an increased accumulation of amino acids in the cell.
CC In this case, an increase in treonine, histidine, glutamic acid and
CC proline is achieved if multiple copies of the gene are transfected into a
CC bacterium. The bacterium used is E. coli.
XX
SO Sequence 588 BP; 113 A; 134 C; 167 G; 174 T; 0 other;

alignment_scores:
Quality: 126.00 Length: 204
Ratio: 1.125 Gaps: 10
Percent Similarity: 54.902 Percent Identity: 24.510

alignment_block:
US-09-466-935-2 x A52690 ..
Align seg 1/1 to: A52690 from: 1 to: 588

8 AlAtyrlLeuLeuThrSerIleIleLeuThrLeuSerProGlySerGlyAl 24
22 GCTTTTGCAGCTTACACCCATATACCGCTATGATGACCGACGAGCAGCA 71
24 aLeaSnThrMetThrThrSerLeuAsnHisGlyTyrProAlaGlyGly 41
72 TATTCCTGCCCTTACCTGCTACGTCAGTCAGTATTT..... 108
41 aLtyrCysTrpAlaSerAspArgThrGlyAspSerTyrCysAlaGlyTyr 57
109 .....CGTCAAGTACCCGCGCGCTGCGGAGGANG 138
58 Arg.....GlyValGlyThrLeuPhe 65
139 ACTCTGGGATTTTGATTTGATGATGTTACTGTGTCGGGCAFTTCATTTTC 188
65 rArgSerValIle.....AlaPheGluValLeuLysTrpAlaGly 79
189 ACTGGCAGTATTGACCCGCGGCGGTACACCTTTGAGTTGGCGGGG 238
79 1aAlaTyrLeuLeuTyrPheLeuGlyIleGlnGlnTrrParGAlaIaGlyAla 95
239 CGGCATATATGTCGTGCTGGCG.....TGGAAATCGCCACCAGC 279
96 1LeaSplLeuYsSerLeuAlaSerThrGlnSerArgArgHisLeuPheG 112
280 CCACCAAGAGAAAGAC.....GGACTTCAGGCAACCAATCAGCTTTTG 323
112 nArgAlaValPheValAsnLeuThrAsnProLysSerIleValPheLeu 129
324 GGCCAGCTTGTGCTTGACGTTGTGAAGCTCAAAATATTTTGTACGCTG 373
374 TTACGGCACTGTCGACGTTGTCTGCCG...CAACACAGCGGTAAAG 420
146 TyrIleValLeuGlyValThrThrIleValAlaSplIleIleValMetI 162
421 TGGGTAGTT...GGCGTACGCTT.....TTGGTGGCGAGAGAT 455
162 eGlyTyrAlaThrLeuAlaGlnArgIleAlaLeuThrPileLysGlyPro 179
456 TGGG.....ACGTTTGGCAAT.....GTGCTGCTGGCGCTGCGGCGG 493
179 ysglMetLysAlaLeuAsnLysIlePheGlySerLeuPheMetLeuVal 195
494 ATCTGTTTCAGCGATTTTTCGCCAGTATGTCGCCAGTTAAATATTCGTG 543
```



```

XX OS Agrobacterium vitis.
XX PN WO200028056-A2.
XX PD 18-MAY-2000.
XX PR 05-NOV-1999; 99WO-US26079.
XX PR 06-NOV-1998; 98US-0107387.
XX PA (CORR ) CORNELL RES FOUND INC.
XX PI Burr TJ, Herlache TC, Zhang H;
XX DR WPI; 2000-376567/32.
XX DR P-PSDB; B11635, B11636, B11637, B11638.
XX PT New protein from Agrobacterium vitis, useful e.g. for imparting
XX PT resistance to disease or stress to plants, is involved in production of
XX PT a hypersensitive response -
XX PS Claim 7; Page 86-87; 157bp; English.
XX PS
XX CC Sequences A61501-A61524 represent nucleotide sequences encoding
XX CC Agrobacterium vitis hypersensitivity response (HR) elicitor proteins
XX CC (B11630-B11688). The HR is a rapid, localized necrosis that is
XX CC associated with the active defence of plants against many pathogens, and
XX CC occurs when a pathogenic organism interacts with a nonhost plant (i.e.
XX CC one in which intracellular bacterial growth and disease development do
XX CC not occur). Like other HR elicitors, the A. vitis elicitor functions in
XX CC non-host plants by causing a rapid hypersensitive response that results
XX CC in walling-off and killing of the pathogen. On grape plants, the A. vitis
XX CC elicitor induces a restricted necrosis of tissues, resulting in the death
XX CC of plant cells and induction of pathogen resistance. A. vitis HR elicitor
XX CC proteins, in non-infectious form, are used to treat plants or their seeds
XX CC to impart resistance to disease, such as those caused by fungi, bacteria
XX CC or viruses; and to enhance growth, e.g., to increase yield or to provide
XX CC earlier germination or maturation. The proteins can also be used to
XX CC control insects, to impart resistance to environmental stresses, e.g.,
XX CC cold, and to improve nutritional value, e.g., altered oil content. The
XX CC same effects can be produced by producing transgenic plants or seeds by
XX CC incorporation of DNA that encodes A. vitis HR elicitor proteins. Use of
XX CC A. vitis HR elicitor proteins, or nucleic acids encoding them, may allow
XX CC control of previously untreatable diseases; provide systemic treatment;
XX CC and eliminate the need for biological control agents or polluting
XX CC chemicals.
XX SQ Sequence 1668 BP; 342 A; 462 C; 423 G; 441 T; 0 other;

```

```

alignment_scores:
  Quality: 123.50      Length: 209
  Ratio: 1.056        Gaps: 6
  Percent Similarity: 55.981  Percent Identity: 20.574

```

alignment_block:

US-09-466-935-2 x A61502 ..

Align seg 1/1 to: A61502 from: 1 to: 1668

```

10 LeuLeuThrSerIle.....IleLeuThrLeuSerProGlySerGlyAl 24
   |||:|||||:|||||:|:|:|:|||||:|||||:|:|:|:|:|:|:|:|:|
970 CTGATTGCTCGATTACCTGCGCTGCGCTTCGACGCCGCCGCGGATTT 1019
   24 aLeuSnhMethrThrSerLeuAsnHISGlyTyrProAlaGly.... 39
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1020 TCCCATGGGTGGCGCATAGTTTCTTTATGCGCGCAAAATCGGGCTGG 1069
   40 .....GlyValTyrCysTrpAlaSerAspArgThrGly 50
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1070 CGGCGTCGATGGCGCATTCGATTCGCTGCG..... 1101

```

```

51 AspSerTyrCysAlaGlyTTPArgGlyValGlyThrLeuPheSerArgse 67
1102 .....TTTCATGATGATCTATGCCATGTCGTCGTCGTCGTCGTCG 1133
   67 rValIle.....AlapheGluValLeuTyrTrpAlaG 78
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1134 GATCATTCGACGATATTTCCCAATATTCGATATTCGATATTCGATAT 1183
   78 lYAlaIaTTrLeuIleTrrPleuGlyIle...GlnGlnTrrPAlaIa 93
   |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1184 GCGCGGCTATCTGCTATCGCGGCTTGGCCACGCGCTTCACCAAGATA 1233
   94 GlYAlaIleAspLeuTyrSerLeuAlaSerThrGlnSerArgTrpHis 110
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1234 CGCGATTCGAAGCAGCTGCTGTCGCCAGCAGACATGATGGCGCGGA 1283
   110 udheGlnArGAlaValAlpHeValaSnLeuThrAsnProLysSerIleValP 127
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1284 MATGATGACGGGTATTCGACCAATGCTTCATCCGMAAGCTGATTT 1333
   127 heLeuAlaIaLeuPheProGlnPheIleMetProGlnGlnProGlnLeu 143
   ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1334 TCGTATCAGTCTCTACACGCAATTCATCGGCAAGACAGCCG...CTG 1380
   144 MetGlnTyrIleValIleuGlyValThrThrIleValAlaSpIleIleVa 160
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1381 TCTCACCAATTCCTTGGGCTGCTTCATCTCCCTGCGCATTTGCTGTG 1430
   160 lMetIleGlyTyrAlaThrLeuAlaGlnArgIleAlaLeuTrpIleLys 177
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1431 GTTCGATCCGTTTGACCTTTTGTCCACCCGCGCATTCGAACTGTG 1480
   177 lYProLysGlnMetLysAlaLeuAsnLysIlePheGlySerLeuPheMet 193
   :|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
1481 TCGTCGTCGCGCAGCGGCTGTTCGAAATCCGATCGGGGCTGTCGCC 1530
   194 LeuValGlyAlaLeuLeuAlaSerAla 202
   1531 TCGCTGGGGCCATTCCTTTTACCGCG 1557
seq_name: /SID52/ycgdata/geneseq/geneseqn/NA2000.DAT:A48443
seq_documentation_block:
ID A48443 standard; DNA; 840 BP.
XX AC A48443:
XX DT 08-SEP-2000 (first entry)
XX DE E. coli L-threonine resistance gene, rhtc.
XX KW L-threonine resistance; L-threonine synthesis; rhtc;
XX KM L-homoserine; L-valine; L-leucine; ds.
XX OS Escherichia coli.
XX XX
XX Key location/Qualifiers
XX PH 187..807
XX FT /*tag= a
XX FT /product= "Rhtc"
XX PI
XX EP1013765-A1.
XX PD 28-JUN-2000.
XX PF 20-DEC-1999; 99EP-0125406.
XX PR 23-DEC-1998; 98RD-0123511.
XX PA (AJIN ) AJINOMOTO KK.
XX PI Livshits VA, Zakatava NP, Aleshin VV, Belareva AV, Tokimakova IL;

```


Quality: 113.00 Length: 209
Ratio: 0.926 Gaps: 10
Percent Similarity: 58.373 Percent Identity: 24.880

alignment_block:

US-09-466-935-2 x A52691 ..

Align seg 1/1 to: A52691 from: 1 to: 636

```
7 PheAlaTYR.....LeuLeuThrSerIleIleLeuTh 17
  |||:|||||
4 TTTCTATTACTTTCAGAGCTTGCACCTTGGGCGGCTATATACCTACC 53
  |||:|||||
17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnH 34
  |||:|||||
54 GCTGGTCACCAAAATGCTTTGGT.....ATGANTCAGGCGCATACGTC 97
  |||:|||||
34 lSGLYTPROAlaGly..GlyValTYrCysTPraLaseRpaGthr 49
  |||:|||||
98 GTCAGTACACATTTATGATTGCTTGTCTTGTCTATCATCAGCATTTGGTC 147
  |||:|||||
50 GlyAspSerTYrCysAlaGlyTPraGlyValGlyThrLeuPheSerAr 66
  |||:|||||
148 .....CTGATTTGCGCGGAGATTTTGGGCGAGCGGCTTTATGATGA 191
  |||:|||||
66 gSerValIleAlaPheGluValLeuYstrPaIaGlyAlaIaTYrLeu 83
  |||:|||||
192 GTCGCGCGTGTGGCTGGCGCTGCACCTGGCGGCGCTTGTCTTGGC 241
  |||:|||||
83 lETTPleuGlyIleGlnGlnTPraGAlaIaGlyAlaIleAspLeuYs 99
  |||:|||||
242 TGTGGTATGTTTTGGGCGCTTTAAACACATGACGATATTTAG 291
  |||:|||||
100 SerLeuAlaSerThr.....GlnSerArgArgHisLeuPheG 112
  |||:|||||
292 ...TTAGCCAGCGCCGAGATCATGAAGCAAGCAGATGAAATATATGCG 338
  |||:|||||
112 nArgAlaValPheValAsnLeuThrAsnPro.....LysSerI 125
  |||:|||||
339 CACCATGTTGGCAGTACGCTGGCTGAATCCGATGTTTACTGATACTT 388
  |||:|||||
125 lValPheValAlaIleuPheProGlnPheIleMetProGlnInPro 141
  |||:|||||
389 TTGTTGATCTGGCAGCTTGGCGGCA.....CTTATGTGGACCA 432
  |||:|||||
142 GlnLeuMetGlnTYrIleValleuGlyValThrThlIleValValAsp 158
  |||:|||||
433 AAA.....CGCTGTTTGACCTGGG.....ACATTAAGCGCTCTTT 470
  |||:|||||
158 eIleValMetIleGlyTYrAlaThrIleuAlaGlnArgIleAlaLeuTP 175
  |||:|||||
471 COTGTGCTTTTGGTCTGCTCTCTTCGAGGCTGGCTGGGACCGCGTC 520
  |||:|||||
175 lElYSGlyProLYSGlnMetLYSAlaLeuAsnLYSlePheGlySerLeu 191
  |||:|||||
521 TGGCGACGCGCAAAAGACACGCGCATTAATCAATCTGTGTGGATGTGT 570
  |||:|||||
192 PheMetLeuValGlyAlaLeuLeuAla 200
  |||:|||||
571 ATGTGCTTATTGCTTCGACGCTGGCG 597
  |||:|||||
```

seq_name: /STD2/gcgdata/geneseq/geneseq/NA197.DAT:T96816

seq_documentation_block:

ID T96816 standard; DNA; 2374 BP.

AC T96816;

DT 12-MAR-1998 (first entry)

DE DNA encoding LysG, LysE and ORF3 from *Corynebacterium glutamicum*.

XX LysG; LysE; ORF3; Lysine transport; regulatory protein; export protein;

```
KW Microbial production; amino acid; animal feed additive; ds.
XX
OS Corynebacterium glutamicum.
XX
FH Key Location/Qualifiers
FT CDS complement (82..954)
FT FT /*tag= a
FT CDS /label= LysG
FT FT 1016..1726
FT FT /*tag= b
FT CDS /label= LysE
FT FT complement (1723..2373)
FT FT /*tag= c
FT FT /label= orf3
XX
PD DEL19548222-A1.
XX
PD 26-JUN-1997.
XX
PF 22-DEC-1995; 95DE-1048222.
XX
PR 22-DEC-1995; 95DE-1048222.
XX
PA (KERJ ) FORSCHUNGSZENTRUM JUELICH GMBH.
XX
PI Eggeling L, Sahm H, Vrljic M;
XX
XX WPI; 1997-333867/31.
XX P-PSDB; W37714-16.
XX
PT Increasing microbial production of amino acids, especially lysine -
PT by improving export carrier activity or corresponding gene
PT expression, also new export and regulatory genes from
PT Corynebacterium
XX
PS Claim 23 and 26; Page -: 16pp; German.
XX
CC This DNA, isolated from Corynebacterium glutamicum, contains the LysG,
CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory
CC protein and an export protein, respectively. Microbial production of
CC amino acids (A) is improved by increasing the export-carrier activity
CC and/or the export gene expression in a microorganism that produces (A).
CC The method is specifically used to increase production of lysine,
CC used as an animal feed additive. Other (A) are variously useful as
CC pharmaceuticals, condiments and intermediates for fine chemicals.
CC This method increases the amount of (A) secreted into the culture medium.
CC Export of (A) has been found to depend on a single gene.
CC NB. This sequence has been created from the information given in table 2
CC of the specification.
XX
SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other:
XX
```

alignment_scores:

Quality: 89.50 Length: 175
Ratio: 1.078 Gaps: 4
Percent Similarity: 47.429 Percent Identity: 20.571

alignment_block:

US-09-466-935-2 x T96816 ..

Align seg 1/1 to: T96816 from: 1 to: 2374

```
3 lEaGlnTYrTPraPheAlaTYrLeuThrSerIleIleLeuThrLeu 19
  |||:|||||
1025 ATGGAATCTTCATACAGCTCTTGTGGGCGCCAGTCTTTACTGTC 1074
  |||:|||||
19 rProGlySerGlyAlaIleAsnThrMetThrThrSerLeu..AsnHisG 35
  |||:|||||
1075 CATCGACGCGCGAATGATCTGTGATTAACAAGCAATTAACCGGGAAG 1124
  |||:|||||
35 lTYrProAlaGlyGlyValTYrCysTPraLaseRpaGthrGlyAsp 51
  |||:|||||
```

```

1125 GACTCATTTGGCGTTCTCTGTCGTTTAAATTCGACGCTTT..... 1168
52 SerTyrCysAlaGlyTrpArgIleValIleGlyThrLeuPheSerArgSerVa 68
   :::: ||||| :::: ||| ::::
1169 TTGTTCATCCGCGGACCTGGCGCTTGATCTTTTCCAAATGCCGCGC 1218
68 IleAlaPheGluValIleuLysTrpAlaGlyAlaAlaTyrLeuIleTrpL 85
   ||| :::: ||| :::: ||| :::: ||| :::: ||| :::: ||| ::::
1219 GATCTGCTCTGATATATATGCGGTGGGTCGATCCGCTTACCTGTTATGCT 1268
85 euGIYIleGlnGlnTrpArgAlaIleGlyAlaIleAspLeuLysSerLeu 101
   :::: :::: :::: |||
1269 TTGCGCGTATGCGACGCAAGAGCGC..... 1294
102 AlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValAlaPheValAs 118
1294 ..... 1294
118 nleuThrAsnProLysSerIleValPheLeuAlaAlaLeuPheProGlnP 135
   :::: ||||| :::: |||||
1295 .ATGACAAACAGAGGTGGAAGCG.....CCACAGA 1322
135 lleMetProGlnGlnProGlnLeuMetGlnTyrIleValLeuGlyVal 151
   :::: ||||| :::: |||||
1323 TCATTGAAGAAACAGAACACCGCTGCCGATGACACGCCCTTGGGCGGT 1372
152 ThrThrIleValIleAlaSpIleIleValMetIleGlyTyrAlaThrLeuAl 168
   :::: ||||| :::: |||||
1373 TCGCGGGTGGCCACTGACACGCGCAACCGGTGGCGGTGAGTGAAGCT 1422
168 aGlnArgIleAlaLeuTrpIleLys 176
   :::: |||||
1423 CGATAACACGCGGTTGGGTAAAG 1447

seq_name: /stds2/gcdata/geneseq/geneseq/NA1998.DAT:V27579
seq_documentation_block:
ID V27579 standard; DNA: 2289 BP.
XX
AC V27579;
XX
DT 26-OCT-1998 (first entry)
XX
DE Human hepatocyte nuclear factor 4 isoform alpha 4 DNA.
XX
KW Hepatocyte nuclear factor 4; HNF4 alpha 4; human;
KM non-insulin-dependent diabetes mellitus; NIDDM; therapy: diagnosis;
XX vaccine; ss.
XX
OS Homo sapiens.
XX
PN W09823780-A1.
XX
PD 04-JUN-1998.
XX
PF 25-NOV-1997; 97WO-US21614.
XX
PR 26-NOV-1996; 96US-0032043.
XX
PA (JOSL-) JOSLIN DIABETES CENT INC.
XX
PI Krolewski AS;
XX
DR WPI: 1998-322753/28.
XX
PT Use of hepatocyte nuclear factor 4 - for developing products for
PT diagnosing and determining risk of diabetes and for identifying
PT agents for treatment of diabetes
XX
PS Claim 11; Fig 3; 43pp; English.
XX
CC This nucleotide sequence encodes human hepatocyte nuclear factor
CC 4 isoform alpha 4 (HNF4 alpha 4). Sequences are also provided for

```

```

CC isoforms alpha 1 (see V27577), alpha 2 (see V27578) and gamma (see
CC V27580). HNF4 is a transcription factor whose ligand has not yet
CC been identified. The isoforms of HNF4 are reported to be derived
CC from alternative splicing events. The alpha isoforms have been
CC mapped to chromosome 20 and the gamma isoform to chromosome 8. A
CC claimed method for determining if an animal is at risk of
CC developing diabetes, comprises evaluating an aspect of HNF4
CC metabolism or structure in the animal, especially as related to
CC HNF4 from chromosome 20. An abnormality in the aspect of HNF4
CC metabolism or structure is diagnostic of being at risk for
CC diabetes, especially non-insulin-dependent diabetes (NIDDM). Also
CC claimed are: (1) a method for evaluating an agent for use in
CC treating diabetes by evaluating the effect of the agent on an
CC aspect of HNF4 metabolism; (2) the agent identified in (1); (3) a
CC method for treating a cell having an abnormality in structure or
CC metabolism of HNF4, comprising administering the agent to the cell;
CC (4) a method for producing a fragment or analogue of HNF4
CC polypeptide having wild-type activity; and (5) a method for
CC producing an HNF4 polypeptide or fragment or analogue having a
CC non-wild type activity.
XX
SQ Sequence 2289 BP; 489 A; 723 C; 653 G; 424 T; 0 other;

```

```

alignment_scores:
Quality: 85.50 Length: 172
Ratio: 1.043 Gaps: 7
Percent Similarity: 47.674 Percent Identity: 22.093

```

```

alignment_block:
US-09-466-935-2 x V27579/rev ..

```

```

Align seg 1/1 to reverse of: V27579 from: 1 to: 2289

```

```

18 LeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnI 34
   ||||| ||||| :::: ::::
1471 CTGTCCCTGGGTGGGCGCCACTGACACATCTGTCCTGAGGTGAG 1422
34 sGlyTyrProAlaGlyValIleTyrCysTrpAlaSerAspArgTrpGlyA 51
   |||
1421 TGGG..... 1418
51 spSerTyrCysAlaGlyTrpArgIleValIleGlyThrLeuPheSerArgSer 67
   :::: ||||| :::: |||||
1417 .....CATGTGTGGCAACGATGACGTGGTCC..... 1388
68 ValIleAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLeuIleTr 84
1387 .....CA 1386
84 pleuGIYIleGlnGlnTrpArgAlaIleGlyAlaIleAspLeuLysSerL 101
   :::: ||||| :::: |||||
1385 TATGTCTCTGATCAGATGAGGTGAGGGGTGGTGGCATGGGGTGCA 1336
101 euAlaSerThrGlnSerArgHisLeuPheGlnArgAlaValAlaPheVal 117
   :::: ||||| :::: |||||
1335 TCGCTGGGGACCTCCGACGACATCTCTGCAACAGTT.....GTC 1292
118 AsnLeuThrAsnProLysSerIleValPheLeuAlaAlaLeu.....Ph 132
   ||||| :::: |||||
1291 AATCTTGGCCATGCGAAGAGCTTGATGATGATGATGATGATGATGAT 1242
132 eProGlnPheIleMet.....ProGlnGlnProGlnLeuMetGln. 145
   :::: ||||| :::: |||||
1241 GCCACGTGATGCTGCAAGGTGGGACACGACGACGACGCTCCCAAG 1192
146 .....TyrIleValLeuGlyValThrThrIleValAlaValAspIle 159
   :::: ||||| :::: |||||
1191 CGGCCACGGAGTCATACGCGGCTTGATGATGATGATGATGATGATGAT 1142
160 ValMetIleGlyTyrAlaThrLeu.....AlaGlnArgIle 171
   :::: ||||| :::: |||||
1141 CTGACACTGGGAACGACGCGCTTGATCTTCCGATGATGATGATGATGAT 1092

```

171 eAlaLeuTriLeuLys 176
 : : : : :
 1091 TGGCATCTGGGTCAA 1076

seq_name: /sids2/gcgdata/geneseq/geneseqn/NA199.DAT:X13048

seq_documentation_block:
 ID X13048 standard; DNA: 6590 BP.

AC X13048;

DT 19-MAR-1999 (first entry)

DE Enterococcus faecalis genome contig SEQ ID NO:111.

KM Enterococcus faecalis: contig: detection: Enterococcal infection;

XX vaccine; attenuation; computer readable medium; ds.

OS Enterococcus faecalis.

XX WO9850555-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08985.

XX 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX (HUMA-) HUMAN GENOME SCI INC.

XX Barash SC, Dillon PJ, Kunsch CA;

DR WPI; 1999-045171/04.

PT New isolated Enterococcus faecalis polynucleotides and polypeptides

PT - used to develop products for the detection of Enterococcus and for

PT use in vaccines for prevention or attenuation of Enterococcus

PS Claim 1; Page 700-703; 2084pp; English.

XX A computer readable medium has been developed which has recorded on it

CC 982 nucleotide sequences isolated from the Enterococcus faecalis genome.

CC X12938 to X13919 represent these nucleotide sequences which are primary

CC nucleotide sequences, also known as contigs. The computer-based system

CC can identify fragments of the Enterococcus faecalis genome with

CC commercial importance. The products can be used to detect the presence

CC of Enterococcus faecalis in samples. They can also be used for

CC diagnosing Enterococcal infection in an animal and monitoring

CC progression of disease, and for identifying agents which can be used to

CC modulate the growth or pathogenicity of Enterococcus faecalis, or

CC another related organism, in vivo or in vitro. In particular the

CC polypeptides encoded by the Enterococcus faecalis nucleotide sequences

CC can be used in vaccines to prevent or attenuate an Enterococcal

CC infection.

SQ Sequence 6590 BP; 1981 A; 1397 C; 1108 G; 2091 T; 13 other;

alignment_scores:

Quality: 85.00

Ratio: 0.859

Percent Similarity: 41.597

alignment_block:

US-09-466-935-2 x X13048/rev ..

Align seg 1/1 to reverse of: X13048 from: 1 to: 6590

7 PheAlaTyrLeuLeu.....ThSerIleIleLeuTh 17

```

5143 TTTGCTTATTATGCTCTTACATCGGATTAACAGTGCACGCTTTTGG 5094
17 rLeuSerProGlySerGlyAlaIle.....AsnThrMetThrThrs 31
5093 TGCACGCGCAGCGTGTGGGNGTATTGCGGACAGTCAATGTTAACAG 5044
31 erLeuAsnHisGlyTyrProAlaGlyGlyValTyrCysTrpAlaSerasp 47
5043 GAATGAATCCGATGCTCCAAATCAGTAATATTTC..... 5009
48 ArgThrGlyAspSerTyrCysAlaGlyTrpArgGlyValGlyThrLeuph 64
5008 ...ACTGGGGAGACATTATCTGCAGACACAGCGGAATT..... 4973
64 eSerArgSerValIleAlaPheGluValLeuLysTrpAlaGlyAlaIat 81
4972 .....ATCGGTATTATT 4960
81 YrLeuIleTrpLeuGlyIleGingIntTrpArgAlaIaGlyAlaIleasp 97
4959 TTGCCGTATGG..... 4949
98 LeuLysSerLeuAlaSerThrGlnSerArgArgHisLeupheGlnArgAl 114
4948 TTATTATCATTTGTTAGAAAACAAATTCATAAATGTGCCGAGATCAT 4899
114 a...ValPheValAsnLeuThrAsnProLysSerIleValPheIleuAla 130
4898 TCATATTATCGTAACGCCAACATTTCACTTTAGTATGTGGCTTAGCAG 4849
130 lAleuPheProGlnPheIleMetPro.....GingInProGlnLeuMetGlnTy 138
4848 CTAATTTTC.....TTAATCATGCGCTGTAGCCGAGCTATTCTTAATGCG 4805
139 .....GingInProGlnLeuMetGlnTy 146
4804 TTAGTTGCTATTATCATGTTGCTTAGAAAAGGTGGCATGTTGCGGG 4755
146 rIleValLeuGlyValThr..... 152
4754 CTTTACTCTTGGTTTAACCTTGTGCCAATGTAATGTTGCTTAAATC 4705
153 .....Thr 153
4704 AAATTTGACGCCAATCCACATTTAAATGATTAATCAACAGGATGACT 4655
154 lIleValIleAspIleIleValMetIleGlyTyrAlaThrLeuAlaGlnAr 170
4654 TTGTTATTACCAATTCGTGCCATGCGTGGGAGCAAGTGGTGACAGC 4605
170 gIleAlaLeuTriPheLysGlyProLysGlnMetLysAlaLeuAsnLysI 187
4604 TTTAGCATTTATGATTCGTTGTAATCAGATAAATAATTAGTTGAATGA 4555
187 lePheGlySerLeu 191
4554 TTAAGGTGCTTGG 4541

```

OM of: US-09-466-935-2 to: Issued_Patents_NA:* out_format : pfs
Date: May 6, 2001 3:40 PM

Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```

@header 11: name: "model"
@Q/cgcn2_1/spofo.spofo/US094663935/runatc_030522001_075709_4814/app.query.fasta_1.5229
-bb=Issued_patents_NA -QPM=fastap -SUFFIX=eni -GAPOP=12.000
-GAPOP=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPRE=0.000
-GAPOP=4.500 -OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-REAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=dlssun62
-TPRAX=human40.cdi.lint-45 -DOCALLGN=200 -THR_SCORE=Pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
USER=US094663935_@cgcn1_1_59 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPRY
-WAIT -THREADS=1

```

Search information block:

Query length: 205

Database sequences: 302621

Search time (sec): 80.320000

score_list:	Strd Orig	ZScore	EScore len	! Documentation
Sequence				

/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-976-254-6 +	184.00	379.86	1.4e-13	720	1	Patena
/cgn2_6/ptodata/2/1na/backfill1.seq:54791933-3 +	188.00	388.64	4.6e-14	720	1	Patena
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-927-074-4 -	82.50	149.00	1.03	1441	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-927-074-4 +	79.50	142.45	2.39	1437	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-083-741-4 +	79.50	142.45	2.39	1437	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-764-233-1 +	79.50	143.33	465.39	49377	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-179-557-16 +	78.00	103.44	2.10	994	1	S
/cgn2_6/ptodata/2/1na/6B.COMB.seq:US-08-927-219-130 +	78.00	116.59	65.78	10014	1	S
/cgn2_6/ptodata/2/1na/6B.COMB.seq:US-08-444-644-27 +	77.50	114.63	84.57	10785	1	S
/cgn2_6/ptodata/2/1na/6A.COMB.seq:US-08-444-644-41 +	77.50	114.56	85.26	10844	1	S
/cgn2_6/ptodata/2/1na/6A.COMB.seq:US-08-444-644-18 +	77.50	113.85	93.41	11528	1	S
/cgn2_6/ptodata/2/1na/6A.COMB.seq:US-08-444-644-32 +	77.50	113.26	100.73	12127	1	S
/cgn2_6/ptodata/2/1na/6B.COMB.seq:US-08-927-219-46 +	77.00	143.15	2.18	844	1	S
/cgn2_6/ptodata/2/1na/6B.COMB.seq:US-08-878-801-3 -	76.50	138.71	3.85	1125	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-453-695-113 +	75.50	127.96	15.28	2347	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-268-161-113 +	75.50	127.96	15.28	2347	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-453-702-113 +	75.50	127.96	15.28	2347	1	S
/cgn2_6/ptodata/2/1na/PCURS.COMB.seq:PCUR-US95-08087-113 +	75.50	127.96	15.28	2347	1	S
/cgn2_6/ptodata/2/1na/6B.COMB.seq:US-09-128-155-17 +	75.50	177.75	9.5e+03	176373	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-210-535-5 +	75.00	118.31	52.71	4899	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-323-443-1 +	74.00	98.46	1.1e+03	31571	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-658-136-2 +	74.00	88.32	2.5e+03	53536	1	S
/cgn2_6/ptodata/2/1na/6A.COMB.seq:US-08-658-136-1 +	74.00	88.31	2.5e+03	53537	1	S
/cgn2_6/ptodata/2/1na/6A.COMB.seq:5169260-3 -	73.50	110.28	14.77	7366	1	Patena
/cgn2_6/ptodata/2/1na/6A.COMB.seq:US-08-822-586-45 +	73.50	106.61	236.31	10095	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-850-880-11 -	73.50	99.15	615.40	19182	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-944-916-11 -	73.50	99.15	615.40	19182	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-944-149-8 +	73.00	114.84	82.28	4557	1	S
/cgn2_6/ptodata/2/1na/6B.COMB.seq:US-09-087-194-34 +	72.00	131.90	9.23	864	1	S
/cgn2_6/ptodata/2/1na/6B.COMB.seq:US-09-088-195-10 +	72.00	117.47	58.72	2889	1	S
/cgn2_6/ptodata/2/1na/6B.COMB.seq:53268270-1 +	72.00	111.73	12.66	4899	1	Patena
/cgn2_6/ptodata/2/1na/6B.COMB.seq:US-09-068-195-21 +	72.00	110.24	148.33	5565	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-356-354-3 +	71.50	114.13	90.12	3625	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-778-656-3 +	71.50	114.13	90.12	3625	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-356-354-1 +	71.50	113.77	94.41	3740	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-778-656-1 +	71.50	113.77	94.41	3740	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-09-131-648-3 +	71.00	122.23	31.88	1643	1	S
/cgn2_6/ptodata/2/1na/5A.COMB.seq:US-08-451-715A-5 -	70.50	115.53	75.36	2662	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-941-445A-4 -	70.50	108.67	181.50	4800	1	S
/cgn2_6/ptodata/2/1na/5B.COMB.seq:US-08-882-083-1 -	70.50	108.12	194.91	5035	1	S

```

;cgml_6/prodata/2/lma/58_COMB.seq:US-08-558-107-1 - 70.50 108.12 194.91 5035
;cgml_6/prodata/2/lma/58_COMB.seq:US-09-243-539-1 - 70.50 108.12 194.91 5035
;cgml_6/prodata/2/lma/58_COMB.seq:US-08-290-878-17 - 70.50 95.24 1.0e+03 15239
;cgml_6/prodata/2/lma/58_COMB.seq:US-09-276-531-19 + 70.00 117.80 56.30 1992
;cgml_6/prodata/2/lma/58_COMB.seq:US-09-276-531-21 + 70.00 117.18 60.36 2101

seq_name: /cgml_6/prodata/2/lma/backfile1.seq:5474933-3

seq_documentation_block:
; Patent No. 5474933
; APPLICANT: EINER, RONALD M.;FUQUA JR., WILLIAM C.
; TITLE OF INVENTION: MARINE MELA GENE
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,945
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 974,837
; FILING DATE: 10-NOV-1992
; APPLICATION NUMBER: 496,804
; FILING DATE: 21-MAR-1990
; SEQ ID NO:3
; LENGTH: 720
5474933-3

Alignment_scores:
Quality: 188.00 Length: 210
Ratio: 1.593 Gaps: 4
Percent Similarity: 56.190 Percent Identity: 28.571

Alignment_block:
US-09-466-935-2 x 5474933-3 ..

Align seg 1/1 to: 5474933-3 from: 1 to: 720

9 tyrleulethrserlleleuthrleusserprogllyalail 25
:::||||| :||| ::::::::::::::::::::|||
122 tvttttaccgaccccttttttgcttcatttacccaggtatgctatgc 171
25 easnthmetthrthrserleuasnhisglytyrrprolaiglyvaltl 42
:::||||| :||| ::::::::::|||
172 actggcgcatagacctctgcgtatgagtacgctgtgc..... 205
42 yrcysttpalaseraspargtthrglyaspsertrycysalaglytprag 58
206 .....ccccgaacc.....ttaggatg 223
59 glyalaglythrleupheserargservallileala..... 70
||||| ||| ::::::::::|||
224 atggttgtgactgacagagcgttgccctctggtgccattgccgccta 273
71 .....pheglualleul 75
||||| :||| ::::::::::|||
274 ggstgtccgcagtatgatgcatactcaacaactcttgatanatttta 323
75 ysttpalaglialaalayrrleuilletripleuglyllegintiprag 91
||||| :||| ::::::::::|||
324 aatggctcggtggcgtctatcttggttacatcgcatTTACATGTCGGCG 373
92 alaaalaglyalalle...aspleuysserleualaSerthrglnsear 107
||| ||| ::::::::::|||
374 gccaaaaggaaantggccaacttgacaaatcctccagctgacTACGTAA 423
107 gartghsluepheglinaarqalavalaphevalasnleuthrasnprolys 124
:||| ||| ::::::: :||| ::::::::::|||
424 tcgaagccctratatcaagaagctttgtcacagccaatgctnaatccaaag 473
124 erllevalpheulaalaleuupheproglinphelietmetproglin 140
:::||||| :||| ::::::::::|||
474 gcrtggccttatgatgtgcctgcctccccttttatcacgctgtaccaa 523
141 proglinleumetlntryllevalleuglyvalthrthrllevalalas 157

```

[illegible]

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-476-254-6

```

seq documentation block:
: Sequence 6, Application US/08476254
: Patent No. 5846531
:
: GENERAL INFORMATION:
: APPLICANT: WEINER, RONALD M.
: APPLICANT: FUQUA, WILLIAM C.
: TITLE OF INVENTION: MARINE MELA GENE
: NUMBER OF SEQUENCES: 11
: CORRESPONDENCE ADDRESS:
: ADDRESS: WATSON COLE STEVENS DAVIS, P.L.L.C.
: STREET: 1400 K STREET NW
: CITY: WASHINGTON
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005-2477
:
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patent in Release #1.0, Version #1.30
:
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/476,254
: FILING DATE: 06-JUN-1995
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: POULOS III, JAMES A.
: REGISTRATION NUMBER: 31,714
: REFERENCE/DOCKET NUMBER: JAP30319C
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 202 628-0088
: TELEFAX: 202 628-8034
: INFORMATION FOR SEQ ID NO: 6:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 720 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHEetical: NO
: ANTI-SENSE: NO
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 167..628
:
: JS-08-476-254-6

```

```
alignment_scores:      Length:    210
                      Quality:     184.00
                        Ratio:       1.546
Percent Similarity:    Gaps:         4
                       Percent Identity: 27.619
```

```
alignment_block:
  ME=09-466-925-3
```

US-09-466-935-2 x US-08-476-254-6

Align seg 1/1 to: US-08-476-254-6 from: 1 to: 720

9 TyrlenuhrserIlelleuThlrLeuSerProlyserGlyalaI 25
 122 TTTTACCGACCTTTTTCGTCTCAATTACACCGATGTGTATGAC 171
 25 eaSnThMeThrThrSerLeuSnHsIGlyTyProAlaGlyGlyAlr 42
 172 ACAGGCCATACCTCGGTATGATCGGTG..... 205
 42 yrcystPrAlaSerAspArgThrGlyAspSerTyrcysAlaGlyTrrpArg 58
 206CGCCGAAC.....TATGATG 223
 59 GlyValGlyThrLeuPheSerArgSerValIleAla..... 70
 224 ATGGTGTGTAGCTACAGCGCTTGCCCTCGGCGATTCGCCCGGTAA 273
 71PheGluValLeu 75
 274 GGGTTCGCCAGTATGATGCTGACATTCACACCTCTCGATATTTAA 323
 75 ysrTrpAlaGlyAlaAlaTyrlLeuIleTrpLeuGlyIleGlnIntrrpArg 91
 324 AATGGGTGGGTGGCTCTATCTGTGGTTACATCGCATTCAGCATGTGGCG 373
 92 AlaAlaGlyAlaIle...AspleuLysSerLeuAlaSerThrGlnSerAr 107
 374 GCCAAGGCAAAATGGCCCAACCTTGACATTCCTCCAGTCAGATGATA 423
 107 gArGHisLeuPheGlnArgAlaValAlaPheValaSnLeuThrAsnProLys 124
 424 TCGAGCGCTAATACCAAGGCTTTGTCAACCGCAATGTCAATCCAAAG 473
 124 exrLeuAlaPheLeuAlaAlaLeuPheProGlnPheIleMetProGlnGln 140
 474 GCGGGCCTTTATGATCTCGCTCCGCCCTTTATTCAGCGGTGACAA 523
 141 ProGlnLeuMetGlnTyrlLeuValleuGlyValThrThrIleValValas 157
 524 GCGATTGCACCAAAATTAATGGTATTCGTCAATTTATTCATGTACCA 573
 157 pIleIleValMetIleGlyTyrlAlaThrLeuAlaGlnArgIleAlaLeuT 174
 574 GTTCTTCACATGCTGTATATGCGGCGG..CGAAACCCTTAAACTGT 622
 174 rPIleLysGlyProLysGlnMetLysAlaLeuAsnLysIlePheGlySer 190
 623 TTTTAAGTCGAGCGCATWACATCAAGTGTATGACCGCATAGAGGAGAT 672
 191 LeuPheMetLeuValGlyAlaLeuLeuAla 200
 673 TTTAATGATCTGTGTGGCTTATGTTGGG 702

seq_name: /cgn2_6/prodata/2/ina/6B_COMB.seq:US-08-927-219-78

```

seq_documentation_block:
: Sequence 78, Application US/08927219
: Patent No. 6187533
:
: GENERAL INFORMATION:
:
: APPLICANT: Bell, Graeme I.
:
: APPLICANT: Yamagata, Kazuya
:
: APPLICANT: Oda, Naohisha
:
: APPLICANT: Kaiseki, Pamela J.
:
: APPLICANT: Furuta, Hiroto
:
: APPLICANT: Horikawa, Yukio
:
: APPLICANT: Menzel, Stephen
:
: TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
:
: TITLE OF INVENTION: GENES HEPATOCTYE NUCLEAR FACTOR (HNF) 1 ALPHA,
:
: TITLE OF INVENTION: AND HNF-4ALPHA
:
: NUMBER OF SEQUENCES: 147
:
: CORRESPONDENCE ADDRESS:
:
: ADDRESSEE: Arnold, White & Durkee
:
: STREET: P.O. Box 4433
:
: CITY: Houston

```



```

STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/927,219
FILING DATE: Concurrently Herewith
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 60/029,679
FILING DATE: 30-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/028,056
FILING DATE: 02-OCT-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 60/025,719
FILING DATE: 10-SEP-1996
ATTORNEY/AGENT INFORMATION:
NAME: Wilson, Mark B.
REGISTRATION NUMBER: 37,259
REFERENCE/DOCKET NUMBER: ARCD:272
TELECOMMUNICATION INFORMATION:
TELEPHONE: 512/418-3000
TELEFAX: 512/474-7577
INFORMATION FOR SEQ ID NO: 78:
SEQUENCE CHARACTERISTICS:
LENGTH: 1441 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
FEATURE:
NAME/KEY: CDS
LOCATION: 20..1414
US-08-927-219-78

```

```

alignment_scores:
Quality: 82.50 Length: 172
Ratio: 1.019 Gaps: 7
Percent Similarity: 47.093 Percent Identity: 22.093

alignment_block:
US-09-466-935-2 x US-08-927-219-78/rev ..
Align seg 1/1 to reverse of: US-08-927-219-78 from: 1 to: 1441

```

```

18 LeuSerProGlySerGlyAlaIleAsnThrMetThrSerLeuAsnH1 34
|||||
1270 CTGTCCTCCGCGGCGGCACATCTGTCGCTGAGGTGAG 1221
34 SGLYTRProAlaGlyGlyValTYrcySTPAlaSerAspArgThrGlyA 51
|||||
1220 TGGG..... 1217
51 spSerTYrcyAlaGlyTRPArgGlyValGlyThrLeuPheSerArgSer 67
:::|||||
1216 .....CATTTGTTGGCAGCATGACGTTGGTTCC..... 1187
68 ValIleAlaPheGluValLeuLysTRPAlaGlyAlaAlaArgLeuIleTr 84
1186 .....CA 1185
84 PleuGlyIleGlnGlnTRPArgAlaAlaGlyAlaIleAspLeuYSerL 101
:::|||||
1184 TATGTTCCATGCATCAGTGAGGCTGCAGGCGGTGGGTGCGGTGCA 1135
101 euAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValPheVal 117
:::|||||
1134 TGGCTGGGGGACCTCCACGACGATCTCTGCAACAGGTT.....GTC 1091

```

```

118 AsnLeuThrAsnProLysSerIleValPheLeuAlaIleLeu.....Ph 132
|||||
1090 AATCTTGCCATGCCGAAAGCGTTGATGAATCTGCTGCATCATCT 1041
132 eProGlnPheIleMet.....ProGlnGlnProGlnLeuMetGln 145
|||
1040 GCCAGGTGATGCTCTGCAGAGTGGGCGACGACGACGACGCTCTCCAAAG 991
146 .....TyrIleValLeuGlyValThrIleValValAspIleIle 159
:::|||||
990 CGGCCACGCGAGTCACTGCGGCTGTGATGATGATCTCCCAACCTCAC 941
160 ValMetIleGlyTyrAlaIleThrLeu.....AlaGlnArgIle 171
:::|||||
940 CTGCACCTGGGAACGACGCCCTGATCTTCCTGATGCTCAGCCCT 891
171 euAlaLeuTRPLeuLys 176
:::|||||
890 TGGCATCTGGGTCCAA 875

```

seq_name: /cgn2_6/plodata/2/ina/5B_COMB.seq:US-08-976-074-4

```

seq_documentation_block:
Sequence 4, Application US/08976074
Patent No. 5840847
GENERAL INFORMATION:
APPLICANT: ABRAMOVITZ, MARK
APPLICANT: GRYGORCZYK, RICHARD
APPLICANT: METTERS, KATHLEEN
APPLICANT: NGUYEN, TRIYEN
APPLICANT: RUSHMORE, THOMAS H.
APPLICANT: SLIPETZ, DEBORAH
TITLE OF INVENTION: DNA Encoding Prostaglandin Receptor FP
NUMBER OF SEQUENCES: 5
CORRESPONDENCE ADDRESS:
ADDRESSEE: J. Mark Hand - Merck & Co., Inc.
STREET: P.O. Box 2000 - 126 E. Lincoln Avenue
CITY: Rahway
STATE: New Jersey
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/976,074
FILING DATE:
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: HAND, J. MARK
REGISTRATION NUMBER: 36,545
REFERENCE/DOCKET NUMBER: 19027DA
TELECOMMUNICATION INFORMATION:
TELEPHONE: (732) 594-3905
TELEFAX: (732) 594-4720
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 1437 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
US-08-976-074-4

```

```

alignment_scores:
Quality: 79.50 Length: 211
Ratio: 0.710 Gaps: 12
Percent Similarity: 53.081 Percent Identity: 22.275

```

alignment_block:
US-09-466-935-2 x US-08-976-074-4 ..

Align seg 1/1 to: US-08-976-074-4 from: 1 to: 1437

```

21 GlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnHisGlyTyrPr 37
   |||||
   :
   :
78 GCGAGCGCGCGCGCGGG.....CGCCATGGCAGCACCC 109
   |||||
   :
37 GAlaGlyValTyrCysTrpAlaSerAspArgThrGlyAspSerTyrC 54
   |||||
   :
110 GAGGCGCTCCGCTTCCTGC.....TCCCAAGAGAGCCCGG 144
   |||||
   :
54 ySalagly..TrpArggly...ValGlyThrLeupheserArgserVal11 69
   |||||
   :
145 CTGGCGGCGCTGGATGACAAAGATCTGTGACTGCATCTCCACAGTTT 194
   |||||
   :
69 eAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLeuIleTrpLeu 86
   |||||
   :
195 GAGAGGAGATGACTGTGAGTGG.....TTGGCTTTATCT 229
   |||||
   :
86 Lylle.....GlnGlnTrpArgAlaGly 94
   |||||
   :
230 CCACAACATGTCATGACATTCACAAACGCTAGTGTCTCTCCAGACT 279
   |||||
   :
95 AlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgHisLeuph 111
   |||||
   :
280 GCGCTTCTTCAACACACACCTGCCAGACGGAACCGGCTTCCGATTT 329
   |||||
   :
111 eGlnArgAlaValPheValAsn.....LeuThrAsnProLysSeri 125
   |||||
   :
330 TTTTTCATGATCTTCATGACAGTGGGATCTGTCAACAGCCCTTCCA 379
   |||||
   :
125 LeValPheLeuAlaIleLeupheProGlnPheIleMetProGlnPro 141
   |||||
   :
380 TCGCATCTTCATGACAGCATATCAGAGATT.....AGACAGAAG 420
   |||||
   :
142 GlnLeuMetGlnTyrIleValLeuGlyValThrIleValValAsp11 158
   |||||
   :
421 TCCAAAGCATCGTTTGTGCTTTGGCCAGCGCGCTGTGTAATCATGATT 470
   |||||
   :
158 e.....IleValMetIleGlyTyrAlaThrL 167
   |||||
   :
471 CTTTGGCCATCTCATCAATGAGCCATAGCAGTATTGTATATCTCTGTG 520
   |||||
   :
167 euAlaGlnArgIleAlaLeuTrpIleLysGlyProLysGlnMetLysAla 183
   |||||
   :
521 ATTAAGAA.....TGGATCCGCTTT...GACCAATCAAAATGTC 555
   |||||
   :
184 LeuAsnLysIlePheGlySerLeuphe.....Me 193
   |||||
   :
556 CTTTGAGCATATTTTGGTATCTGCATGTGTCTGTCTGTGTGCCACT 605
   |||||
   :
193 tLeuValGlyAlaLeuLeuAlaSerAlaArg 203
   |||||
   :
606 TCTTCTAGCGAGTGTGTGATGGCCATTGAGCGG 636
   |||||
   :
seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-083-741-4

seq_documentation_block:
; Sequence 4, Application US/08083741
; Patent No. 5869281
; GENERAL INFORMATION:
; APPLICANT: ABRAMOWITZ, MARK
; APPLICANT: GRYGORCZYK, RICHARD
; APPLICANT: METTERS, KATHLEEN
; APPLICANT: NGUYEN, TROYEN
; APPLICANT: RUSHMORE, THOMAS H.
; APPLICANT: SLIPETZ, DEBORAH
; TITLE OF INVENTION: DNA Encoding Prostaglandin Receptor FP
; NUMBER OF SEQUENCES: 5
; CORRESPONDENCE ADDRESSES:
; ADDRESSEE: John W. Wallen, III

```

```

; STREET: P. O. Box 2000, 126 E. Lincoln Avenue
; CITY: Rahway
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07065
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/083,741
; FILING DATE: 25-JUN-1993
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: WALLEN, JOHN W.
; REGISTRATION NUMBER: 35,403
; REFERENCE/DOCKET NUMBER: 19027
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908)594-3905
; TELEFAX: (908)594-4720
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1437 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
; US-08-083-741-4

```

alignment_scores:
Quality: 79.50 Length: 211
Ratio: 0.710 Gaps: 12
Percent Similarity: 53.081 Percent Identity: 22.275

alignment_block:
US-09-466-935-2 x US-08-083-741-4 ..

Align seg 1/1 to: US-08-083-741-4 from: 1 to: 1437

```

21 GlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnHisGlyTyrPr 37
   |||||
   :
   :
78 GCGAGCGCGCGCGCGGG.....CGCCATGGCAGCACCC 109
   |||||
   :
37 GAlaGlyValTyrCysTrpAlaSerAspArgThrGlyAspSerTyrC 54
   |||||
   :
110 GAGGCGCTCCGCTTCCTGC.....TCCCAAGAGAGCCCGG 144
   |||||
   :
54 ySalagly..TrpArggly...ValGlyThrLeupheserArgserVal11 69
   |||||
   :
145 CTGGCGGCGCTGGATGACAAAGATCTGTGACTGCATCTCCACAGTTT 194
   |||||
   :
69 eAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLeuIleTrpLeu 86
   |||||
   :
195 GAGAGGAGATGACTGTGAGTGG.....TTGGCTTTATCT 229
   |||||
   :
86 Lylle.....GlnGlnTrpArgAlaGly 94
   |||||
   :
230 CCACAACATGTCATGACATTCACAAACGCTAGTGTCTCTCCAGACT 279
   |||||
   :
95 AlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgHisLeuph 111
   |||||
   :
280 GCGCTTCTTCAACACACACCTGCCAGACGGAACCGGCTTCCGATTT 329
   |||||
   :
330 TTTTTCATGATCTTCATGACAGTGGGATCTGTCAACAGCCCTTCCA 379
   |||||
   :
111 eGlnArgAlaValPheValAsn.....LeuThrAsnProLysSeri 125
   |||||
   :
330 TTTTTCATGATCTTCATGACAGTGGGATCTGTCAACAGCCCTTCCA 379
   |||||
   :
125 LeValPheLeuAlaIleLeupheProGlnPheIleMetProGlnPro 141
   |||||
   :
380 TCGCATCTTCATGAAAGCATATCAGAGATT.....AGACAGAAG 420
   |||||
   :

```

```

142 GlnLeuMetClnTrrIleValLeuGlyValThrPheIleValAlaSpl1 158
    :::::::::::::::::::::
421 TCCAGGACATCGTTCTGCTTTGGCCAGCGGCTGTGATCATCATGATT 470
158 e::::::::::::::::::::::::::IleValMetIleGlyTrrAlaThrL 167
    :::::::::::::::::::::
471 CTTTGGCCATCTCATCATGAGCCATAGACGATTTGTATGCTTTCG 520
167 euAlaGlnArgIleAlaLeuTrrPheLeuGlyProLysGlnMetLysAla 183
    :::::::::::::::::::::
521 ATAAGAA.....TGGATCCGGCTT...GACCAATCAATGTC 555
184 LeuAsnLysIlePheGlySerLeuPhe.....Me 193
    :::::::::::::::::::::
556 CTTTGACGATTTTGTGATCTGATCATGCTGTTTCTGCTGTGCCACT 605
193 tLeuValGlyAlaLeuLeuAlaSerAlaArg 203
606 TCTTCTAGGACGATGTGATGCGCATTTGAGCGG 636

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-764-233A-1
seq_documentation_block:
; Sequence 1, Application US/08764233A
; Patent No. 5716849
; GENERAL INFORMATION:
; APPLICANT: Ligon, James M.
; APPLICANT: Schupp, Thomas
; APPLICANT: Beck, James J.
; APPLICANT: Hill, Dwight S.
; APPLICANT: Neft, Snezana
; APPLICANT: Ryals, John A.
; TITLE OF INVENTION: Genes For The Biosynthesis Of Soraphen
; NUMBER OF SEQUENCES: 10
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Ciba-Geigy Corporation
; STREET: 520 White Plains Road, P.O. Box 2005
; CITY: Tarrytown
; STATE: NY
; COUNTRY: USA
; ZIP: 10591
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/764,233A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/729,214
; FILING DATE: 09-OCT-1996
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/258,261
; FILING DATE: 08-JUN-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: Meigs, J. Timothy
; REGISTRATION NUMBER: 38,241
; REFERENCE/DOCKET NUMBER: 1506/CIP6
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (919) 541-8587
; TELEFAX: (919) 541-8689
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 49377 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:
; ORGANISM: Sorangium cellulosum
; IMMEDIATE SOURCE:

```

```

; CLONE: p98/L, pJL3, and pVKM15
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 383..760
; OTHER INFORMATION: /product= "SorB"
; OTHER INFORMATION: /note= "This gene encodes a protein that is highly homolog
; OTHER INFORMATION: the reductase domains of type I PKs such as eryA from
; OTHER INFORMATION: Saccharopolyspora erythraea."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 927..19874
; OTHER INFORMATION: /product= "Sora"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs t
; OTHER INFORMATION: are known to be involved in the synthesis of polyketide
; OTHER INFORMATION: compounds."
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 942..7115
; OTHER INFORMATION: /product= "Module 1 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 7203..12884
; OTHER INFORMATION: /product= "Module 2 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 13455..19616
; OTHER INFORMATION: /product= "Module 3 of Sora"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19871..46318
; OTHER INFORMATION: /product= "Sorb"
; OTHER INFORMATION: /note= "Gene product is highly homologous to type I PKs ge
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 19870..24556
; OTHER INFORMATION: /product= "Module 1 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 24638..30820
; OTHER INFORMATION: /product= "Module 2 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 30881..35446
; OTHER INFORMATION: /product= "Module 3 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 35528..40114
; OTHER INFORMATION: /product= "Module 4 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 40190..46318
; OTHER INFORMATION: /product= "Module 5 of SorB"
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: 46851..47891
; OTHER INFORMATION: /product= "Sora"
; OTHER INFORMATION: /note= "The protein encoded by the sorB gene is highly
; OTHER INFORMATION: homologous to the methyltransferase from Streptomyces
; OTHER INFORMATION: hygrosopicus that is involved in the synthesis of the
; OTHER INFORMATION: polyketide rapamycin."
; US-08-764-233A-1

```

```

alignment_scores:
    Quality: 79.50      Length: 103
    Ratio: 1.445        Gaps: 5
    Percent Similarity: 53.398      Percent Identity: 32.039

```

```

alignment_block:
US-09-466-935-2 x US-08-764-233A-1 ..

```

```

Align seq 1/1 to: US-08-764-233A-1 from: 1 to: 49377

```



```

;
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; FRAGMENT TYPE: N-terminal
; IMMEDIATE SOURCE:
; CLONE: pAH4808
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1..10844
; OTHER INFORMATION: /note= "Function = "Expression
; OTHER INFORMATION: Vector Coding Sequence"
US-08-444-644-41

alignment_scores:
    Quality: 77.50      Length: 169
    Ratio: 0.969        Gaps: 9
    Percent Similarity: 47.337    Percent Identity: 25.444

alignment_block:
US-09-466-935-2 x US-08-444-644-41 ..

Align seg 1/1 to: US-08-444-644-41 from: 1 to: 10844

21 GlySerGlyAlaIleAsnThrMetThrSer..... 31
||| :|||||
8378 GGAATGCCCGCAGCATTAATACCGTTGAGAGAGCGGGCGAGCCGCCAGG 8427
32 ....LeuAsnHisGlyTyrProAlaGlyGlyValTyrCysTrp..... 44
||| :|||||
8428 AATATACACCACTGACAGTACCGGACCGGTGACCTGTCGACACGCA 8477
45 .....AlaSerAspArgThrGlyAspSerTyrCysAlaGlyTyr 57
|||| :|||||
8478 CGCGTGGCTGGTTTCATCATCTACGCGCGCGAGTTTTCGCGAGTGTG 8527
58 ArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheGluValLe 74
|||| :|||||
8528 AAG...GTTTCAATATCTTTACTGCCACCGCCATCGCC..... 8563
74 ulysTrpAlaGlyAlaIleAlaTyrLeuIleTyrPheGlyIleGlnGlnTrpA 91
8563 ..... 8563

91 rGAlaAlaGlyAlaIleAspLeuLysSerLeuAlaSerThrGlnSerArg 107
||| :|||||
8564 ..TGTTTAGCTGCTGCTCAGGCGTTCGCGGCGGCGGAGTCTCTCT 8611
108 ..ArgHisLeuPheGlnArgAlaValAlaPheValAsnLeuThrAspProLys 123
||||| :|||||
8612 GCAGACACCTT...CAGCGC.....GTAACCGT 8637
124 SerIleValPheLeuAlaIleAlaLeuPheProGlnPheIleMet..ProGln 140
||||| :|||||
8638 GGTATTATCAAACTTCGCGCTGTAATCCGACGCGCTCATCCGCGCGT. 8686
140 InProGlnLeuMetGlnTyrIleValLeuGlyValIleThrIleValVal 156
||||| :|||||
8687 .....GCTTTCAGCTTATCGAGAAATVCGTTA 8713
157 AspIleIleValMetIleGlyTyrAlaThrLeuAlaGlnArgIleValLe 173
||||| :|||||
8714 ACAGTGGCGGATATGCTTTTCAGAGCGGAAATCCCGCGCCCTTACAG 8763
173 uTrp 174
|||
8764 CTGG 8767

seq_name: /cgn2_6/prodata/2/ina/6A_COMB.seq:US-08-444-644-18

seq_documentation_block:
; Sequence 18, Application US/08444644
; Patent No. 601555
; GENERAL INFORMATION:
```

```

;
; APPLICANT: Fiden, Phillip M.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR SPECIFIC
; TITLE OF INVENTION: ANTIBODY-NEUROPHARMACEUTICAL OR DIAGNOSTIC AGENT
; TITLE OF INVENTION: CONJUGATES
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,644
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/232,246
; FILING DATE: 07-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,458
; FILING DATE: 26-NOV-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/05077
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/404,089
; FILING DATE: 07-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Wagner, Richard W.
; REGISTRATION NUMBER: 34,480
; REFERENCE/DOCKET NUMBER: ALK88-15AAAZ
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 18:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 11528 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; HYPOTHEITICAL: NO
; ANTI-SENSE: NO
; IMMEDIATE SOURCE:
; CLONE: pAH4602
; FEATURE:
; NAME/KEY: misc-feature
; LOCATION: 1..11528
; OTHER INFORMATION: /note= "Function="Expression Vector
; OTHER INFORMATION: Coding Sequence"
US-08-444-644-18

alignment_scores:
    Quality: 77.50      Length: 169
    Ratio: 0.969        Gaps: 9
    Percent Similarity: 47.337    Percent Identity: 25.444

alignment_block:
US-09-466-935-2 x US-08-444-644-18 ..

Align seg 1/1 to: US-08-444-644-18 from: 1 to: 11528

21 GlySerGlyAlaIleAsnThrMetThrSer..... 31
||| :|||||
9064 GGAATGCCCGCAGCATTAATACCGTTGAGAGAGCGGGCGAGCCGCCAGG 9113
```

```

32 ....LeuSnhISGIYrProAlaGlyValTYrCYsTrp..... 44
      |||
      :||| ||| ||| |||
9114 AATATACAACTGAGCTGAGCGGCTGAGCTGAGCA 9163
45 .....AlaSerSparGThrglyAspSerTYrCYsAlaGlyTrp 57
      :||| ||| ||| ||| ||| ||| ||| |||
9164 CGCTGCGTGGCTTCTACATCTACCGGCGGAGTTTGGCGGTGTG 9213
58 ArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheGluValLe 74
      :||| ||| ||| ||| ||| ||| ||| |||
9214 AAG...GTTCAATATCTTCTTACGCCACCGCCATCCGC..... 9249
74 uLysTrpAlaGlyAlaValTYrLeuIleTrpLeuGlyIleGlnGlnTrpA 91
9249 ..... 9249
91 rGAlaAlaGlyAlaIleAspLeuYSerLeuAlaSerThrGlnSerArg 107
      :||| ||| ||| ||| ||| ||| ||| |||
9250 ..TGTTTAGCTGCTGCTCAGGCGTTCGCGGCGGCGGATCTCTCT 9297
108 ..ArgHISLeuPheGlnArgAlaValPheValAsnLeuThrAsnProLys 123
      ||| ||| ||| ||| ||| ||| ||| |||
9298 GCAGACACCTT...CAGGC.....GTAACCGT 9323
124 SerIleValPheIleAlaAlaLeuPheProGlnPheIleMet..ProGln 140
      :||| ||| ||| ||| ||| ||| ||| |||
9324 GGTTTTCAAACTTCGCGCTGTATCCGCGAGGCGCTCATCCGCGGT. 9372
140 InProGlnLeuMetGlnTYrIleValLeuGlyAlaThrTrpIleValVal 156
      :||| ||| ||| ||| ||| ||| ||| |||
9373 .....GCTTCAGTTATCGAAGATATCGTTA 9399
157 AspIleIleValMetIleGlyTYrAlaThrLeuAlaGlnArgIleAlaLe 173
      :||| ||| ||| ||| ||| ||| ||| |||
9400 ACAGTGGCGGTAATGCTTCAGAGGCGGAATGCGCGGCTTAAACAG 9449
173 uTrp 174
      |||
9450 CTGG 9453
seq_name: /cgn2_6/ptodata/2/lna/6A_COMB.seq:US-08-444-644-32
seq_documentation_block:
; Sequence 32, Application US/0844644
; Patent No. 601555
; GENERAL INFORMATION:
; APPLICANT: Fiden, Phillip M.
; TITLE OF INVENTION: TRANSFERRIN RECEPTOR SPECIFIC
; TITLE OF INVENTION: ANTIBODY-NEUROPHARMACEUTICAL OR DIAGNOSTIC AGENT
; NUMBER OF SEQUENCES: 46
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
; STREET: Two Militia Drive
; CITY: Lexington
; STATE: MA
; COUNTRY: USA
; ZIP: 02173
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/444,644
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/232,246
; FILING DATE: 07-JUL-1994
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/800,458
; FILING DATE: 26-NOV-1991

```

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: PCT/US90/05077
; FILING DATE: 07-SEP-1990
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/404,089
; FILING DATE: 07-SEP-1989
; ATTORNEY/AGENT INFORMATION:
; NAME: Wagner, Richard W.
; REGISTRATION NUMBER: 34,480
; REFERENCE/DOCKET NUMBER: ALK88-15AAAZ
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (617) 861-6240
; TELEFAX: (617) 861-9540
; INFORMATION FOR SEQ ID NO: 32:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 12127 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: circular
; MOLECULE TYPE: DNA (genomic)
; IMMEDIATE SOURCE:
; CLONE: pAH4807
; FEATURE:
; NAME/KEY: misc.feature
; LOCATION: 1..12127
; OTHER INFORMATION: /note= "Expression
; OTHER INFORMATION: Vector Coding Sequence"
US-08-444-644-32

```

```

alignment_scores:
      Quality: 77.50      Length: 169
      Ratio: 0.969      Gaps: 9
      Percent Similarity: 47.337      Percent Identity: 25.444
alignment_block:
US-09-466-935-2 x US-08-444-644-32 ..

```

Align seg 1/1 to: US-08-444-644-32 from: 1 to: 12127

```

21 GlySerGlyAlaIleAsnThrMetThrSer..... 31
      ||| :||| ||| ||| ||| ||| ||| |||
7602 GGAGTGGCGGCAATATATACCGTTGAGAGAGCGGCGGAGCCGAGG 7651
32 ....LeuSnhISGIYrProAlaGlyValTYrCYsTrp..... 44
      |||
      :||| ||| ||| ||| ||| ||| ||| |||
7652 AATATACAACTGAGCTGAGCGGCGGTGACCTGCGGCAACGCA 7701
45 .....AlaSerSparGThrglyAspSerTYrCYsAlaGlyTrp 57
      :||| ||| ||| ||| ||| ||| ||| |||
7702 CGCTGCGTGGCTTCTACATCTACCGGCGGCGGCGGCGGATCTCTCT 7751
58 ArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheGluValLe 74
      :||| ||| ||| ||| ||| ||| ||| |||
7752 AAG...GTTCAATATCTTCTTACGCCACCGCCATCCGC..... 7787
74 uLysTrpAlaGlyAlaValTYrLeuIleTrpLeuGlyIleGlnGlnTrpA 91
7787 ..... 7787
91 rGAlaAlaGlyAlaIleAspLeuYSerLeuAlaSerThrGlnSerArg 107
      :||| ||| ||| ||| ||| ||| ||| |||
7788 ..TGTTTAGCTGCTGCTCAGGCGTTCGCGGCGGCGGCGGATCTCTCT 7835
108 ..ArgHISLeuPheGlnArgAlaValPheValAsnLeuThrAsnProLys 123
      ||| ||| ||| ||| ||| ||| ||| |||
7836 GCAGACACCTT...CAGGC.....GTAACCGT 7861
124 SerIleValPheIleAlaAlaLeuPheProGlnPheIleMet..ProGln 140
      :||| ||| ||| ||| ||| ||| ||| |||
7862 GGTTTTCAAACTTCGCGCTGTATCCGCGAGGCGCTCATCCGCGGT. 7910
140 InProGlnLeuMetGlnTYrIleValLeuGlyAlaThrTrpIleValVal 156

```



```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(1122)
US-08-878-801-3
```

```
alignment_scores:
    Quality: 76.50      Length: 209
    Ratio: 0.900        Gaps: 12
    Percent Similarity: 40.670    Percent Identity: 23.445
```

alignment_block:

US-09-466-935-2 x US-08-878-801-3/rev ..

Align seg 1/1 to reverse of: US-08-878-801-3 from: 1 to: 1125

```
5 TTTTTP.....PheAlaTyrLeuThrSer11 14
|||||
568 TGGTGGCATGCGACTGCGCAGCAGCTCTTGGCGAGTGGGATGATGCTG 519
|||||
14 eilleuThrLeuSerProGlySerGlyAlaIleAsnThrMetThrPhe 31
|||||
518 TCCTCTGATATGCGCTCCAGT.....GTGACAG 490
|||||
31 erleuAsnHisGlyTyrProAlaGlyValTyrCysTrpAlaSerasp 47
:::|::|
489 GTAAATACACCGCGGAGTCCAGAAAGTGGAATTCACGCTTCGCTGAC 440
|||||
48 .....ArgThrGlyaspSerTyr..... 53
|||||
439 ATGCACGATGCCCGCGCTCCGCCACAGAGTACTGATGCGACATGATAT 390
|||||
54 .....CysAlaGlyTrpArgGlyValGlyThrLeuPheSera 66
|||||
389 GGCTTCTCGAATGTGCTCACTTATAGGGGCTCTGTGTCATG..... 350
|||||
66 rgservAlIleAlaPheGluValLeuLysTrp..... 76
|||||
349 .....TCAGTACGGCTGGCGTCTGCTGCTGCTGACG 320
|||||
77 .....AlaGlyAlaAlaTyrLeuIleTyrLeuG1 86
|||||
319 GCCTGCTGAAGGGAGATCTGCACGCGCTCATCGCATATGATG..... 275
|||||
86 ylieGlnGlnTrpArg.....AlaAlaGly 94
|||||
274 .....CCTGCATGAGACGAAAGATGTTCTGTAGATGACAGCCGAAAGC 229
|||||
95 AlaIleAspLeuLysSerLeuAlaSerThrGlnSerArgAlaGHisLeuPhe 111
|||||
228 TCGTGGGTCTCTCTCGAGTACGACACCGCATGATGCGCATCTGCT 179
|||||
111 eGlnIleAlaValPheValAsnLeuThrAsnPro.....LysSerIle 126
|||||
178 TGAATGACCTACTCTCCGCTCTCAGCAGGCGCCACAGCAGGAGTTTC 129
|||||
126 alphaLeuAlaIleAlaLeuPheProGlnPheIleMetProGlnGlnProGln 142
|||||
128 AATTCTCGCGCTCTGTTT.....TTTCTGTCCAAACCAATCT... 88
|||||
143 LeuMetGlnTyrIleValLeuGlyValThrThrIleValAlaAspIleI1 159
|||||
87 .....GTTGATCTCTCT 77
|||||
159 eValMetIleGlyTyrAlaThrLeu 167
|||||
76 GGTGATCTCTGGCGCAGTCTTCTC 52
|||||
```

seq_name: /cgn2_6/ptodata/2/1na/5A_COMB.seq:US-08-453-695A-113

seq_documentation_block:

; Sequence 113, Application US/08453695A

; Patent No. 5708143

; GENERAL INFORMATION:

; APPLICANT: Suzuki, Shintaro

; TITLE OF INVENTION: Protocadherin Materials and Methods

; NUMBER OF SEQUENCES: 115

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Marshall, O'Toole, Gerstein, Murray, &

; STREET: 233 South Wacker, 6300 Sears Tower

; CITY: Chicago

; STATE: Illinois

; COUNTRY: USA

; ZIP: 60606

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/453,695A

; FILING DATE:

; CLASSIFICATION: 530

; ATTORNEY/AGENT INFORMATION:

; NAME: No. 5708143and, Greta E.

; REGISTRATION NUMBER: 35,302

; REFERENCE/DOCKET NUMBER: 32658

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 312/474-6300

; TELEFAX: 312/474-0448

; TELEX: 25-3856

; INFORMATION FOR SEQ ID NO: 113:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 2347 base pairs

; TYPE: nucleic acid

; STRANDEDNESS: single

; TOPOLOGY: linear

; MOLECULE TYPE: cDNA

US-08-453-695A-113

```
alignment_scores:
    Quality: 75.50      Length: 76
    Ratio: 1.606        Gaps: 4
    Percent Similarity: 61.842    Percent Identity: 32.895
```

alignment_block:

US-09-466-935-2 x US-08-453-695A-113 ..

Align seg 1/1 to: US-08-453-695A-113 from: 1 to: 2347

```
5 TTTTTPheAlaTyrLeu.....LeuThrSerIleIleLeuThrle 18
|||||
1189 TGGTGGTGGCAGTGGCTGTGCTGCTGCTGCTGCTGCTGCTGCTGCTAT 1238
|||||
18 user: ProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnHis 34
|||||
1239 GTCTCTCTAGCTCTCAGGCT.....TCAGCCTGGCAGAA 1273
|||||
35 GlyTyrProAlaGlyValTyrCysTrpAlaSerAspArgThrGlyAs 51
|||||
1274 GTCCTGCCAGCTCCAGGGCTCCAAAGGTGATGTGCTCTGCACTTCAT 1323
|||||
51 pSerTyrCysAlaGlyTrpArgGlyValGly.....ThrLeuPhe 65
|||||
1324 CACATTTGTGGCATCGACGGGTACAGGCTTTTCTACAAACCTATTCT 1373
|||||
65 eArgSerValIleAlaPheGluVal 73
|||||
1374 CATGAATCTGCTCATTTCAGGCTC 1399
|||||
```

9

1

2

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:34:19 ; Search time 26.05 Seconds
(without alignments)
540.814 Million cell updates/sec

Title: US-09-466-935-2

Perfect score: 1061

Sequence: 1 MTEWMFAYLITSTILLSP.....KIRGSLFMLVALLASARHA 205

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 198801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

PIR 67:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	686	64.7	138	2	A65187
2	459.5	43.3	205	2	G82358
3	389	36.7	209	2	C82990
4	228	21.5	241	2	C75329
5	195.5	18.4	212	2	F64940
6	195	18.4	210	2	F69975
7	193	18.2	210	2	G83082
8	192.5	18.1	207	2	E83703
9	190	17.9	212	2	G82200
10	186	17.5	208	2	C82471
11	180.5	17.0	207	2	D83187
12	180.5	17.0	216	2	F83051
13	178.5	16.8	213	2	C82523
14	178.5	16.8	216	2	C55580
15	166	15.6	206	2	C82139
16	156	14.7	204	2	B83729
17	155	14.6	153	2	S23222
18	154.5	14.6	208	2	G84086
19	150	14.1	204	2	F83306
20	147	13.9	206	2	S76178
21	146	13.8	213	2	F83444
22	143.5	13.5	204	2	B82410
23	141	13.3	211	2	T36373
24	132	12.4	209	2	E82388
25	127.5	12.0	222	2	F82353
26	126.5	11.9	212	2	T43921
27	126.5	11.9	223	2	H64759
28	126	11.9	195	2	A65036
29	124	11.7	210	2	D84016

30	123.5	11.6	197	2	B83280	hypothetical prote
31	116	10.9	203	2	T30317	chemotaxis protein
32	113	10.7	211	1	Q0EC5A	hypothetical 23k p
33	109	10.3	205	2	B83356	conserved hypothet
34	101.5	9.6	4588	2	T28667	dyein beta heavy
35	94	8.9	457	2	B64769	proline transport
36	92	8.7	449	2	F83328	probable sodium/al
37	89.5	8.4	366	2	F83970	stage V sporulatio
38	88	8.3	204	2	B82166	conserved hypothet
39	87.5	8.2	465	2	A83046	probable amino aci
40	87	8.2	210	2	F64115	hypothetical prote
41	84.5	8.0	200	2	D83100	probable transport
42	84	7.9	190	2	T03459	hypothetical prote
43	84	7.9	297	2	H83888	transmembrane lipo
44	83	7.8	674	2	E82954	conserved hypothet
45	82.5	7.8	149	2	C83056	hypothetical prote

ALIGNMENTS

```
RESULT 1
A65187
hypothetical 15.4 kd protein in recQ-PlDB intergenic region - Escherichia coli (strai
N:Alternat names: hypothetical protein fl38
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: A65187; S30714
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: A65187
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-138 <BIAT>
A:Cross-references: GB:AE000458; GB:U00096; NID:g2367299; PIDN:AACT6827.1; PID:g17902
A:Experimental source: Strain K-12, Substrain MG1655
R:Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992
A:Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.
A:Reference number: S30660; MUID:92358234
A:Accession: S30714
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'V', 2-138 <DAN>
A:Cross-references: EMBL:M87049
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992
C:Genetics:
A:Gene: y1gk

Query Match 64.7%: Score 686; DB 2; Length 138;
Best Local Similarity 99.3%: Pred. No. 5.7e-58;
Matches 137; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 68 VIAPFVLKAGAAAYLIWIGIOQWRAAGAIIDLKSLASTOSRRHLFORAVNLTPNKSIVF 127
: |||||
DB 1 MIAFVLKAGAAAYLIWIGIOQWRAAGAIIDLKSLASTOSRRHLFORAVNLTPNKSIVF 60
: |||||

QY 128 IAALEPQFIMPOQPOLMOYIVLGVTTIVDIIWIGVITLAQRIALMTKGRKOKALKMKI 187
: |||||
DB 61 IAALEPQFIMPOQPOLMOYIVLGVTTIVDIIWIGVITLAQRIALMTKGRKOKALKMKI 120
: |||||

QY 188 FGSLEFMYGALLASARHA 205
: |||||
DB 121 FGSLEFMYGALLASARHA 138
: |||||

RESULT 2
G82358
conserved hypothetical protein VC0136 [imported] - Vibrio cholerae (group O1 strain N
```

C:Species: *Vibrio cholerae*
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 17-Nov-2000
C:Accession: G62358
R:Haideleberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Giffin, M.L.; Dodson, R.J.;
Chaidson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F.
L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:title: DNA Sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*.
A:Reference number: A82035; MUID:20406833
A:Accession: G62358
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-205 <HE>
A:Cross-references: GB:AE004104; GB:AE003852; NID:99654534; PIDD:AAFP9313.1; GSDB:GN001
A:Experimental source: serogroup O1; strain N16561; biotype El Tor
C:Genetics:
A:Gene: VC0136
A:Map position: 1
C:Superfamily: hypothetical protein b1798

```

Query Match          43.3%; Score 459.5; DB 2; Length 205;
Best Local Similarity 43.4%; Pred. No. 2.5e-36;
Matches 89; Conservative 44; Mismatches 69; Indels 3; Gaps 2;

OY 1 MTLKMFAYLLTSLITLSPGSAINTMTTSLNNG--YPAGGYCMAKSDRTGDSYCAQMR 58
Dd 1 MDIHVLAFLYLLTAAVFASTAPSGSTVNSISNGLSYGTRISLGAIIIGLIGLACHIVLG-I 59
OY 59 GVGTLFSSRVIAFEVLKWAAGAYLIIMLGIOQRAAGAIIDKSLASTOSRHLFORAVEYN 118
Dd 60 GIGALVNASALAFILIKKIGTAAYLVWGLGICWKRDRAPLITATTTSHESQALLKRAVLIN 119
OY 119 LTNPKSIVLALFPQFIMPOQPOLMOYIVLGVTTIVDIIMVIGYATLQRIALMTKGP 178
Dd 120 LTNPKSIVLALFPQFIDPRDHMPQFLVIGITTVITDAIVMGYATLAAQLGRITRSP 179
OY 179 KQKRLNKIFGSLTMLVGLLASAR 203
Dd 180 NIMTRMKNLFGSMFGCGMLLATAR 204

RESULT 3
C82990
hypothetical protein PA5249 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C82990
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mircoguchi, S.D.; Warrenner, P.; Hickey, M.V.; Br
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lam,
J.; Lory, S.; Olson, M.V.
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: C82990
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-209 <STO>
A:Cross-references: GB:AE004937; GB:AE004091; NID:q9951553; PIDN:AA608634.1; GSPDB:GN001
C:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5249

Query Match          36.7%; Score 389; DB 2; Length 209;
Best Local Similarity 36.4%; Pred. No. 1.2e-29;
Matches 83; Conservative 37; Mismatches 64; Indels 44; Gaps 4;

OY 1 MTLKMFAYLLTSLITLSPGSAINTMTTSLNNGYPAGGYCMAKSDRTGDSYCAQMR-- 58
Dd 1 MLVSTWFAFLACMAAISTSPGAGIASHMSCGLQYGFARG-----YNAL 44
OY 59 -----GVGTLFSSRVIAFEVLKWAAGAYLIIMLGIOQRAAGAIIDKSLAS 103

```

Db 45 GUGGAGAGUAGVALLATSAATSLIKWEGVATLVLANQWAP-----PQALS 99
 104 TQSRHHLEQ-----RAVEVNLTNKRSIVFLTALEPQFIMPOQPOLMOMYIVLGVTITVVD 157
 Db 100 TDGERPLRPLTLVRLGFLVNAANSNKRAVIFMLAVLPQFIDPHQPLLAQYLLMGSTMIVVD 159
 Qy 158 ITVMGATLTQALTAIWKSPKQKALKIKGSLFMVLGALLASRHA 205
 Db 160 ILVAGYGLAARVLRVLRSPQQLVNTRTASLFLVGAGGLATVTRRA 207

RESULT 4
C755329
conserved hypothetical protein - *Deinococcus radiodurans* (strain RI)
C:Species: *Deinococcus radiodurans*
C:Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 17-Mar-2000
C:Accession: C753329
R:White, O.; Elsen, J.A.; Heidelberg, J.P.; Hickey, E.K.; Peterson, J.D.; Dodson, R.U.;
S.; Shen, M.; Vamathevan, J.J.; Lam, P.; McDonald, L.; Uterback, T.; Zalewski, C.;
S.; Smith, H.O.; Venter, J.C.; Fraser, C.M.
Science 286, 1571-1577, 1999
A:Title: Genome sequence of the radioresistant bacterium *Deinococcus radiodurans* RI.
A:Reference number: A75250; MUID:20036896
A:Accession: C753329
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-241 <WHI>
A:Cross-references: GB:AE002038; GB:AE000513; NID:g6459790; PIDN:AAF11548.1; PID:g64545
A:Experimental source: strain RI
C:Genetics:
A:Gene: DR1999
A:Map position: 1

```

Query Match      21.5%; Score 228; DB 2; Length 241;
Best Local Similarity 30.38; Pred. No. 2.0e-14;
Matches 64; Conservative 41; Mismatches 84; Indels 22; Gaps 6;

QY      8 AYLTSILTLTSPSGAINTMTWTSLNHCYPAAGVYCWASDRPDGSYCAGWR-----G 59
Db      38 AFIYAAYVALLLPFGILMTILARSLGGR-----WAGIOSALCTGAGGVHYLASAVG 90
          |:::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
QY      60 VGLTFSRSVIAFEYLVKMGAAVYLIWLGIOQWRAAGAILDK--SLASTQSR--HLPQRA 114
          :::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      91 LSAIIMASSLAFSVYKKAAGAAVLIYLIGIRVLISKEALSKKEASLAAAPERSQRLFTQG 150
QY      115 VFVWLTNPDKSLVFLAALPEPFIHQPOLM-QYVLGVTTIVVDLIYIGVATLAQRIAL 173
Db      151 AMTELNLNKTLAFLAAYIPQVFKPATGTHVQGFLLLTGTSVYVNTLNMALVATLAGEFLGA 210
QY      174 WIKG-PKOMKALNKFGLSFLMLVGLALLASAR 203
          ::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
Db      211 RLQGNPRFGQKVASGGAMTALCTYVAVER 241

RESULT      5
F64940
hypothetical protein b1798 - Escherichia coli (strain K-12)
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 29-Sep-1999
C:Accession: F64940
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: F64940
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-212 <BLAT>
A:Cross-references: GB:AE000274; GB:U00096; NID:g1788089; PIDN:AACT4868.1; PID:g17880
A:Experimental source: strain K-12, substrain MG1655

```

C;Superfamily: hypothetical protein b17998

Query Match	18.4%;	Score 195.5;	DB 2;	Length 212;
Best Local Similarity	28.6%;	Pred. No. 2.7e-11;		
Matches 61;	Conservative 45;	Mismatches 82;	Indels 25;	Gaps 10;

QY	3	LEWPAVLLIIIIILSLGSGAI----	NEMTSLNHGPAQ-GYYCMA	SDTGTG	SYC--	A	55
		: : : : : : :					
Db	8	LNIV-TYLVGAIFIVLVP	GPNTLFLV	KNSSVSGMKG	VLACGVF-----	IGAVIMFL	60
		: : : : : : :					
QY	56	GMRGVLTFSRSVIAFEV	ELKMACAAVILMLG	ITOO	RAGAIDLK---	SLASTQSRH--	110
		: : : : : :					
Db	61	AMAGVATLTKPTPLIF	FNIVRYLGAFLY	LLYLSK	LIITAY----	LKGKNS	EASDEPQ
		: : : : : : :					
QY	111	FORAVEVNLNPKSIV	FLAALFPQFIM	QOPOL-MQY	VLGTTIVYDI	IVMIGAT	169
		: : : : : : :					
Db	117	EKRALIISLTNP	KAILFVYSEF	VOEIDV	NAPHTGIS	FLATLIELV	5FCLSP
		: : : : : : :					
QY	170	RIALMIKQPKOM-K	ALNKIFGSLF	MLV	GALLAS		201
		: : : : : : :					
Db	177	FVTOYIRTKK	LAIVGNSLIG	IMEV	GFAP	RLAT	209
		: : : : : : :					

RESULT 6
F69975 dihydrodipicolinate reductase homolog yrhP - *Bacillus subtilis*

C:Date: 05-Dec-1997 #sequence_revision 05-Dec-1997 #text_change 21-Jul-2000
C:Accession: F69975
R:Kunst, F.; Ogasawara, N.; Moszer, I.; Albertini, A.M.; Alloni, G.; Azevedo, V.; Bertone
A.; Biron, S.; Broiliet, S.; Bruschi, C.V.; Caldwell, B.; Capuano, V.; Carter, N.M.; Chien
C.; Ehrlich, S.D.; Emmerson, P.T.; Enli, K.D.; Errington, J.; Fabret, C.; Ferrari, E.
Nature 390, 249-256, 1997
A:Authors: Foulger, D.; Fritz, C.; Fujita, M.; Fujita, Y.; Fuma, S.; Galizzi, A.; Gallen
lech, J.; Harwood, C.R.; Henaut, A.; Hilbert, H.; Holappel, S.; Hosono, S.; Hullo, S.;
Koester, P.; Konigstein, G.; Krogi, S.; Kumano, M.; Kurita, K.; Ladiou, A.; Ladinios,
A:Authors: Labber, J.; Lazarevic, V.; Lee, S.M.; Levine, A.; Liu, H.; Masuda, S.; Meisel
Y, M.; Ogawa, K.; Ogizawa, A.; Oudega, B.; Park, S.H.; Parro, V.; Pohl, T.M.; Portele
Rieger, M.; Rivolta, C.; Rocha, E.; Roche, B.; Rose, M.; Sadaie, Y.; Sato, T.; Scanlon
A:Authors: Schelch, S.; Schroeter, R.; Scifone, F.; Sekiguchi, J.; Sekowska, A.; Seron
Kneuchl, M.; Tamakoshi, A.; Tanaka, T.; Tepsuta, P.; Tognoni, A.; Tosio, V.; Uchiyama
T.; Winters, P.; Wipit, A.; Yamamoto, H.; Yamane, K.; Yasumoto, K.; Yata, K.; Yoshida, K
A:Authors: Yoshikawa, H.F.; Zumbstein, E.; Yoshikawa, E.; Dancho, A.
A:Title: The complete genome sequence of the Gram-positive bacterium *Bacillus subtilis*.
A:Reference number: A69580; MUID:98044033

Query Match	18.4%;	Score 195;	DB 2;	Length 210;
Best Local Similarity	27.6%;	Pred. No. 3e-11;		
Matches	56;	Conservative	39;	Mismatches 88;
				Indels 20;
				Gaps 6

QY 8 AYLLSIIILTLTSPGSGAINTMTTSLNHGYPAGGYCMASDRIGDSCACGR-----GVGT 62
 Db 6 AYIPIDAMVILIPGADMTVMKNTLRYG--PKAGRYNIIIGLATGLSF--WTVAIIIGLSV 61
 QY 63 LESRSVIAFEVLKMGAAVLINLIGIQOMRAAGAIIDL-----KLSLTSQSR--RLHFOR 113
 Db 62 VIASKVILFETTKIKYGAALVIYGVKSPFAKSMFLSDMQSOAKNNASSPKRYTKTSFMQ 121
 QY 114 AEFVNLINPKSIVFLAALFEPQFIMPQQPOLMQYIVLGVTTVIIVDIIVMIGATLAQRIAL 173
 Db 122 GSLSLNINPKYIVLVVYTMPQFININGNINQOOLLIIASITLTILAVLWFLVLYIINYARK 181

```
QY 174 WIKGPQOMKALNKIEGSLFMLVG 196
    1:1 1 1 1 1 1 1 1 1 1 1
Db 182 WMKNSKFQKVFKQKITG--IILVG 202
```

RESULT 7
G83082

C:Species: *Pseudomonas aeruginosa*
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: 683082
R:Stover, C.K.; Yam, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warriner, P.; Hickey, M.J.; L
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lorry, S.; Olson, M.V.
N:Date 406, 955-964, 2000
A:Title: Complete genome sequence of *Pseudomonas aeruginosa* PA01, an opportunistic pat
hogen
;Reference number: M01D:20437337

Query Match	18.28;	Score 193;	DB 2;	Length 210;
Best Local Similarity	28.38;	Pred. No. 4.6e-11;		
Matches 58; Conservative	31;	Mismatches 106;	Indels 10;	Gaps 6

Qy 8 AYLTSLITLSPSGAINTTTSLNHYPGAGYCC--MASDRGDSYCAGMREYGLFS 65
 Db 5 AFLLAAVYLVYVPGEDMLLEFQTGARGORRRALVTALGLARACHVLMAA-TGLALLR 63
 Qy 66 RSVAIEFVILKMGAAVYLITLGLTQOMRAAG-AIDUKSLAST---QSRHLLFORAVFNLT 120
 Db 64 TAPWTFEDLVRLGAAVYLAMLGLQMLRGGLALPTSDAGSAVYVPHNARRALLRGLLTNLL 123
 Qy 121 NPKSIVFLAALFQFIMFQPOL-MQYIVLGVTIIVVDIIMIGATYLAORIALIMLWG-P 178
 Db 124 NPKALLFCSEVLLPQVSPSEAGSLAVQFALGLTVLVLGLAFDCAYALLAGRLGWLASRP 183
 Qy 179 KQKRALNKIFGSLFEMLVGALLASAR 203
 Db 184 RAQRLOQWGFGLLIGFVRLALLR 208

RESULT 8
 EB3703
 hypothetical protein BH0429 [imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
 C:Accession: EB3703
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A>Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a
 A:Reference number: A83650; MUID:20263314
 A:Accession: EB3703
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-207 <STO>
 A:Cross-references: GB:AP001508; GE:BA000004; NID:g10172890; PIDN:BB04148.1; GSPDB:G
 A:Experimental source: strain C-125
 C:Genetics:
 A:Gene: BH0429

Query Match	18.1%;	Score 192.5;	DB 2;	Length 207;
Best Local Similarity	26.3%;	Pred. No. 5e-11;		
Matches 57; Conservative	50;	Mismatches 87;	Indels 23;	Gaps 6;

```
OY 1 MLEWMPAYLLTSLILSPSGAINTMTSLNHGYPAGGYCMASDRTGDSYCAGR-- 58
| | : : : | | | : : : | | | : : : | | |
Db 1 MDLTSLSLFAGVAVLLTMGPDPILVLAQSMQNRQAGV-----TALGCTGILVH 53
OY 59 -----GVGLTFSRSVIAFEVLKMGAAVLLWLGIOQWRAAG---AIDLKSLASTOSRRH 109
| | : : : | | | : : : | | | : : : | | |
Db 54 ITAAVVGISALIIYQSALAFVYVYACAAVLLYLAAMKAFQEKGGSLIDKO---TTILA YGA 110
OY 110 LFORAVFVNLTNPKSIYFLAALFPQFIMPOQPOL-MQYIVLGVTIVVDIIVMIGVATLA 168
| | : : : | | | : : : | | | : : : | | |
Db 111 LYKKGIMVNLNPKSVLFLFLLALPQFVNSGAGSAPMOMLLLGVFLLQAFILPSVSWFA 170
OY 169 QRIA-LMIGKPKOMKALNKIFGSLFMLVGCALLASARH 204
| | : : : | | | : : : | | | : : : | | |
Db 171 EKVGLLMRSSPFIKNOMHRIKGLLALIGLVAFSKN 207

RESULT 9
G82200
conserved hypothetical protein VC1421 [imported] - Vibrio cholerae (group O1 strain N169
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: G82200
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
A:Accession: G82200
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-212 <HEI>
A:Cross-references: GB:AE004221; GB:AE003852; NID:99655918; PIDN:AAF94578.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC1421
A:Map position: 1

Query Match 17.9%; Score 190; DB 2; Length 212;
Best Local Similarity 27.4%; Pred. No. 8.9e-11;
Matches 58; Conservative 35; Mismatches 99; Indels 20; Gaps 5;

OY 1 MLEWMPAYLLTSLILSPSGAINTMTSLNHGYPAGGYCMASDRTGDSYCAGW--- 57
| | : : : | | | : : : | | | : : : | | |
Db 1 MYIONFEAFFIITLITLPGDLTALVIRNTSRAGFADG---C---TSLGICGGLFVH 53
OY 58 -----RGVGTLSRSVIAFEVLKMGAAVLLWLGIOQWRA---AGAILDKSLASTOSRR 108
| | : : : | | | : : : | | | : : : | | |
Db 54 AFPSAIGISAILIAQSAEFLQIVKMGAAVLLWLGISLSRLMKTKGQIEVASIAHAQFRL 113
OY 109 HLFORAVFY-NLTNPKSIYFLAALFPQFIMPOQPOLMQRIVLGVTIVVDIIVMIGVATL 167
| | : : : | | | : : : | | | : : : | | |
Db 114 TSLRGFSLNLTNPKAVFYLAFLPQFINTPYSPLAQSLLMALHFAIAMWQCGGLAGA 173
OY 168 AQRIALMIGKPKOMKALNKIFGSLFMLVGCALL 199
| | : : : | | | : : : | | | : : : | | |
Db 174 LSSAKNLKKNASFMRMMEGTGVLVALGIKL 205

RESULT 10
G82471
conserved hypothetical protein VCA0355 [imported] - Vibrio cholerae (group O1 strain N16
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: G82471
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, F
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: AB2035; MUID:20406833
```

```
A:Accession: G82471
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-208 <HEI>
A:Cross-references: GB:AE004372; GB:AE003853; NID:99657741; PIDN:AAF96263.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0355
A:Map position: 2

Query Match 17.5%; Score 186; DB 2; Length 208;
Best Local Similarity 29.3%; Pred. No. 2.1e-10;
Matches 63; Conservative 35; Mismatches 89; Indels 28; Gaps 9;

OY 1 MLEWMPAYLLTSLILSPSGAINTMTSLNHGYPAG-GYYCMASDRTGDSY-- 53
| | : : : | | | : : : | | | : : : | | |
Db 1 MDLNTSLFLIVACLAINTMGPDPVIVTSNTMKGLVTGKRAMGL-----GVGYEV 52
OY 54 --CAGMRGVGTLSRSVIAFEVLKMGAAVLLWLGIO---QWRAGAIDLKSLASTOSR 107
| | : : : | | | : : : | | | : : : | | |
Db 53 HTLLASLGLSATILSSAVAFSAVKMIGAAVLYLVGQSLISMWR--GGSITKVSSEVED 110
OY 108 RHLFORAVFVNLTNPKSIYFLAALFPQFIMPOQPOL-MQYIVLGVTIVVDIIVMIGVAT 166
| | : : : | | | : : : | | | : : : | | |
Db 111 KNVFOGVIVSVLNPKEVALFLFSLFQFIDTSSGASMKOLWLGILFVLTWCNIIYAS 170
OY 167 LAQRIALMIGKPKOMKALNKIFGSLFMLVGCALLAS 201
| | : : : | | | : : : | | | : : : | | |
Db 171 VGSWFSRPNRSORYSRGLGVSGLV--LIG--LAS 201
```

```
RESULT 11
D83187
hypothetical protein PA3665 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: D83187
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.;
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: AB2950; MUID:20437337
A:Accession: D83187
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-207 <STO>
A:Cross-references: GB:AE004786; GB:AE004091; NID:99949822; PIDN:AA607053.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA3665
C:Superfamily: hypothetical protein b1798

Query Match 17.0%; Score 180.5; DB 2; Length 207;
Best Local Similarity 26.5%; Pred. No. 6.9e-10;
Matches 57; Conservative 34; Mismatches 97; Indels 27; Gaps 6;
```

```
OY 6 WFAVLLTSLILSPSGAINTMTSLNHGYP-----AGCVTCMASDRTGDSYCAGW---- 57
| | : : : | | | : : : | | | : : : | | |
Db 5 WALFVPACFALNLAPGPNMLTSLNNAARHGFAVTASLAGGRLLA-----FAGMLALA 56
OY 58 -RGVGTLSRSVIAFEVLKMGAAVLLWLGIOQWRAAGAILDKSLASTOSRR-----HLF 111
| | : : : | | | : : : | | | : : : | | |
Db 57 ASGLALVLTHTSWLPLAIKLGAAVLLWLAVOLWRT---DAQPLANESAPRPSILVRG 112
OY 112 QRAVAVNLTNPKSIYFLAALFPQFIMPOQPOLMQRIVLGVTIVVDIIVMIGVATLAORI 171
| | : : : | | | : : : | | | : : : | | |
Db 113 RQEFIVAAQGNPRAILIFAPLFPQFIDPQPLGAGPAGQAGAAFLLEWLAIALYVAGLHL 172
OY 172 ALMIGKPKOMKALNKIFGSLFMLVGCALLASARHA 205
| | : : : | | | : : : | | | : : : | | |
```


Db 173 GRLLAGORARLRLNRCALLGSLGLLSRRPA 207

RESULT 12

conserved hypothetical protein PA4757 [imported] - Pseudomonas aeruginosa (strain PA01)

C:Species: Pseudomonas aeruginosa

C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

C:Accession: F83051

R:Stover, C.K.; Plam, X.O.; Erwin, A.L.; Micozuchi, S.D.; Warren, P.; Hickey, M.J.; Badman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, .; Lory, S.; Olson, M.V.

Nature 406, 959-964, 2000

A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen

A:Reference number: A82950; MUID:20437337

A:Accession: F83051

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-216 <STO>

A:Cross-references: GB:AE004889; GB:AE004091; NID:99951014; PIDN:AA608143.1; GSPDB:GN001

A:Experimental source: strain PA01

C:Genetics: PA4757

A:Gene: PA4757

C:Superfamily: hypothetical protein b1798

Query Match 17.0%; Score 180.5; DB 2; Length 216;

Best Local Similarity 28.0%; Pred. No. 7.2e-10;

Matches 60; Conservative 41; Mismatches 90; Indels 23; Gaps 9;

Qy 4 EWMFAVLLTSLITLSPGSAINTMTTSLNHGYPAG-----GYCMAASDRGDS--YCAG 56

Db 8 DFW-ITVLGVVIVLILPGNSLFLVLAISAQGVATGRACGVF-----LGDVLMLLS 60

Qy 57 WRGVGTLFSRSVIAFEVLKMGAAVLIWLGIQO---WRAAGAILDKSLASTOSRHL-F 111

Db 61 ALGVASILKAEPMLFGLKYLGAAYLFLYLGVMRLGAMRKLRNEATRAQAQEDVHQPF 120

Qy 112 QRAVAVNTNKSIVFLAALPPQIMP--QOPOLMQYIVLGVTIVDIIMIGYATFLA 169

Db 121 RKALLLSLNSPKAILFLFISFIOQVDPGYAVPGV-SFLVLAIVLEIVSALVLSFLIPTGV 179

Qy 170 RIALMIKGPOMKA-LNKIFGSLFMLVGLALLASA 202

Db 180 RLAMFRKRQRLAAGATSGVALVGVGVKLATA 213

RESULT 13

C82523

amino acid transporter XF2730 [imported] - Xylella fastidiosa (strain 9a5c)

C:Species: Xylella fastidiosa

C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 20-Aug-2000

C:Accession: C82523

R:anonymous, The Xylella fastidiosa Consortium of the Organization for Nucleotide Sequence

Nature 406, 151-157, 2000

A:Title: The genome sequence of the plant pathogen Xylella fastidiosa.

A:Reference number: A82515; MUID:20355717

A>Note: for a complete list of authors see reference number A59328 below

A:Accession: C82523

A:Molecule type: DNA

A:Status: preliminary

A:Residues: 1-213 <STM>

A:Cross-references: GB:AE004079; GB:AE003849; NID:99107960; PIDN:AAF85515.1; GSPDB:GN001

A:Experimental source: strain 9a5c

R:Stimpson, A.J.G.; Reinech, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.; Brites, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carrier, H. as-Melo, E.; Docena, C.; El-Dorri, H.; Facincani, A.P.; Ferreira, A.J.S.

submitted to Genbank, June 2000

A:Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Frohm J.D.; Junqueira, M.L.; Kemper, E.L.; Kitaajima, J.P.; Krieger, J.E.; Kuramae, E.E.; Laigt Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins, F. A:Authors: Martins, E.M.F.; Matsuuma, A.Y.; Menck, C.F.M.; Miracca, E.C.; Miyaki, C.Y.; F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri, D.A.

Rodrigues, V.; Rosa, A.U. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa A:Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv M.; Tsubako, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.

A:Reference number: A59328

A:Contents: annotation

C:Genetics:

A:Gene: XF2730

Query Match 16.8%; Score 178.5; DB 2; Length 213;

Best Local Similarity 25.0%; Pred. No. 1.1e-09;

Matches 51; Conservative 38; Mismatches 94; Indels 21; Gaps 4;

Qy 9 YLITSITLSPGSAINTMTTSLNHGYPAG-----GYCMAASDRGDSYGCAGWG 59

Db 21 FILTVLMFISIPGAMMFVLOQSGKGVKGLAAVLGTEIGVEIVY-----LITALG 72

Qy 60 VGTFLFSRSVIAFEVLKMGAAVLIWLGIQOORAAAGAILDKSLASTOSRHLFORAVFVL 119

Db 73 ISVTLKEPDSIYGLGIGAAVLIYAVLSWPRONASNOTPPASRSSYGTFTIGVLIWL 132

Qy 120 TNPKSIVFLAALPPQIMPQOPOLMQYIVLGVTIVDIIMIGYATFLAQRALMIKGP 179

Db 133 TNKIVLFLSLIPQV-PRDSNMTFMVYGLIFNTSGILVNSVALLADRVNRL---R 188

Qy 180 QMKALKIFGSLFMLVGLALLASAR 203

Db 189 SVTFMVFVPIIFLSIVAVLSMKAR 212

RESULT 14

C55580

hypothetical protein (cara 3' region) - Pseudomonas aeruginosa

C:Species: Pseudomonas aeruginosa

C>Date: 25-Aug-1995 #sequence_revision 25-Aug-1995 #text_change 29-Sep-1999

C:Accession: C55580

R:Kwon, D.H.; Lu, C.D.; Walthall, D.A.; Brown, T.M.; Houghton, J.E.; Abdelal, A.T.

J. Bacteriol. 176, 2532-2542, 1994

A:Title: Structure and regulation of the cara operon in Pseudomonas aeruginosa and P

A:Reference number: A55580; MUID:9422830

A:Accession: C55580

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-216 <KWO>

A:Cross-references: GB:U04992; NID:9451649; PIDN:AAA19047.1; PID:9451652

C:Superfamily: hypothetical protein b1798

Query Match 16.8%; Score 178.5; DB 2; Length 216;

Best Local Similarity 27.9%; Pred. No. 1.1e-09;

Matches 60; Conservative 42; Mismatches 88; Indels 25; Gaps 9;

Qy 4 EWMFAVLLTSLITLSPGSAINTMTTSLNHGYPAG-----GYCMAASDRGDS--YCAG 56

Db 8 DFW-ITVLGVVIVLILPGNSLFLVLAISAQGVATGRACGVF-----LGDVLMLLS 60

Qy 57 WRGVGTLFSRSVIAFEVLKMGAAVLIWLGIQOORAAAGAILDKSLASTOSR-----RHL 110

Db 61 ALGVASILKAEPMLFGLKYLGAAYLFLYLGVMRLGAMR-KLRNEATRAQAQEDVHQPF 119

Qy 111 FORAVAVNTNKSIVFLAALPPQIMP--QOPOLMQYIVLGVTIVDIIMIGYATFLA 168

Db 120 FROALLLSLNSPKAILFLFISFIOQVDPGYAVPGV-SFLVLAIVLEIVSALVLSFLIPTGV 178

Qy 169 ORIALMIKGPOMKA-LNKIFGSLFMLVGLALLASA 202

Db 179 VRLAMFRKRQRLAAGATSGVALVGVGVKLATA 213

RESULT 15

C82139

conserved hypothetical protein VC1939 [imported] - Vibrio cholerae (group O1 strain N

C:Species: Vibrio cholerae

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:38:09 ; Search time 16.79 Seconds
(without alignments)
418.247 Million cell updates/sec

Title: US-09-466-935-2

Perfect score: 1061
Sequence: 1 MTELEWFAVLLTSLITLSP.....KIRGSLFMLVGLALASARRA 205

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 3425486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SWISSProt_39:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	898.5	84.7	206	1 RHTB_ECOLI	P27847 escherichia
2	195.5	18.4	212	1 YEAS_ECOLI	P76249 escherichia
3	195	18.4	210	1 YRHP_BACSU	O05406 bacillus su
4	180.5	17.0	216	1 YBFT_PSEAE	P38102 pseudomonas
5	147	13.9	206	1 YG27_SYNI3	P74343 synechocyst
6	126.5	11.9	223	1 YAHN_ECOLI	P75693 escherichia
7	126	11.9	195	1 YFIR_ECOLI	P38101 escherichia
8	122	11.5	206	1 RHTC_ECOLI	P27846 escherichia
9	116	10.9	203	1 CHPE_PSEAE	O87005 pseudomonas
10	113	10.7	211	1 YGCA_ECOLI	P11667 escherichia
11	111.5	10.5	206	1 YGGA_AERSA	P70775 aeromonas s
12	94	8.9	457	1 PROY_ECOLI	P77327 escherichia
13	92	8.7	456	1 PROY_SALTY	P37460 salmonella
14	89.5	8.4	236	1 LYSE_CORGL	P94633 corynobacte
15	87	8.2	210	1 YD07_HAEIN	O57320 haemophilus
16	81	7.6	732	1 ATZN_ECOLI	P37617 escherichia
17	79.5	7.5	384	1 CYNA_ECOLI	P17583 escherichia
18	79.5	7.5	589	1 FUR4_SCHPO	Q10279 schizosach
19	79	7.4	579	1 YN05_CAEEL	Q03614 caenorhabdi
20	78.5	7.4	366	1 SPSE_BACSU	P07373 bacillus su
21	78.5	7.4	471	1 PROX_MYXXA	P56601 myxococcus
22	78	7.4	1694	1 SN_MOUSE	O62230 mus musculu
23	77	7.3	225	1 YGGA_AERHY	P52047 aeromonas h
24	77	7.3	348	1 OPSP_SARXA	P79914 sarcocentro
25	77	7.3	543	1 SGUT_VIBPA	P96169 vibrio para
26	75.5	7.1	261	1 SP3T_BACSU	O01625 bacillus su
27	75.5	7.1	445	1 IF36_HUMAN	O64725 homo sapien
28	75.5	7.1	460	1 ARCD_ECOLI	P77429 escherichia
29	75.5	7.1	566	1 KHT2_KJULA	P53387 kluyveromyc
30	75	7.1	373	1 Y3BK_THETE	P30715 thermoprote
31	75	7.1	440	1 YJ99_MYCTU	Q10858 mycobacteri
32	74.5	7.0	394	1 FT5W_HAEIN	P45064 haemophilus
33	74.5	7.0	460	1 NU4M_ONCMY	P11631 oncoirynchu

34	74.5	7.0	504	1 OCLN_CHICK	Q91049 gallus gall
35	74.5	7.0	564	1 BMRP_CANAL	P28873 candida alb
36	74.5	7.0	1411	1 Y297_HUMAN	O15040 homo sapien
37	74.5	7.0	2145	1 U520_CAEEL	O90290 caenorhabdi
38	74	7.0	204	1 YAGU_ECOLI	P77262 escherichia
39	74	7.0	351	1 OPSP_SARDI	P79898 sarcocentro
40	73.5	6.9	153	1 YJEM_SALTY	Q92311 salmonella
41	73.5	6.9	276	1 NOSY_PSEST	P19845 pseudomonas
42	73.5	6.9	317	1 Y302_MYCGE	P47544 mycoplasma
43	73.5	6.9	461	1 YIFK_ECOLI	P27837 escherichia
44	73	6.9	362	1 CKRA_MOUSE	O91121 mus musculu
45	73	6.9	459	1 AAPA_BACSU	O06005 bacillus su

ALIGNMENTS

```

RESULT 1
RHTB_ECOLI
ID RHTB_ECOLI STANDARD; PRT; 206 AA.
AC P27847;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMOSERINE/HOMOSERINE LACTONE EFFLUX PROTEIN.
GN RHTB.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=92338234; Pubmed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
from 84.5 to 86.5 minutes."
RL Science 257:771-778(1992).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JUN-1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=93313167; Pubmed=10386596;
RA Zakateva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
RL "The novel transmembrane Escherichia coli proteins involved in the
amino acid efflux."
RT FEBS Lett. 452:228-232(1999).
RN [4]
RP FUNCTION: CONDUCTS THE EFFLUX OF HOMOSERINE AND HOMOSERINE
LACTONE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO
BE INTRODUCED IN POSITIONS 60 TO EXTEND THIS ORF.
-----
This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL Outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
-----
EMBL: M87049; AAA67620.1; ALT_FRAME.
EMBL: AE000458; AAC76827.1; ALT_FRAME.
PIR: S30714; S30714.
Ecogene: EG11469; rhtB.
InterPro: IPR001123; -.
Pfam: PF01810; lyse; 1.
Transport: Transmembrane.
TRANSMEM 5 25 POTENTIAL.

```


CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: Z99117; CAB14652.1; -
DR EMBL: U93874; AAB80873.1; -
DR Subtilist: BG12304; yrbp.
DR InterPro: IPR001123; -
DR Pfam: PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 75 95 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
SQ SEQUENCE 210 AA; 23389 MW; 2A37D9419FDB0A58 CRC64;

Query Match 18.4%; Score 195; DB 1; Length 210;
Best Local Similarity 27.6%; Pred. No. 2.3e-11;
Matches 56; Conservative 39; Mismatches 88; Indels 20; Gaps 6;

QY 8 AYLTSTIITLSPGCAINTMTSLNHGYPAGVYCMASDRFGDGYCAGNR-----GVGT 62
DB 6 AYPIAMWVITPGADTLMVKNKNTLRIG-PAKGRYNIIGLGTLSF---WYVIALIGLSV 61
QY 63 LFSRSYIAFEVLKMGAAVILWLGIOQMRAGAIDL-----KSLASTOSR--RHLFOR 113
DB 62 VIASKVILFTTITKYLGAALVILYGVKSPFAKMFSLDDMQSAKMASSPKRYITSTFQ 121
QY 114 AVEFVNTLNKRSIVLEALFPQIFMPQPOLMOYIVLGVTIVDIIWIGVTTAQRAL 173
DB 122 GSLSNILNKRKTIVLVYVTIMPOFINLNGINNOOILILASITLLAVLWFLVYIIDYAKK 181
QY 174 WIKPKROMKALNKIFGSLFMWG 196
DB 182 WMKNSKFKQFKYKRTIG--IILVG 202

RESULT 4
YB7_PSEAE STANDARD; PRT; 216 AA.
AC P38102; Q9H4V5; -
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN PA4757.
GN PA4757.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01.
RX MEDLINE=94222830; Pubmed=8169201;
RA Kwon D.-H., Lu C.-D., Walthall D.A., Brown T.M., Houghton J.E.,
RA Abdelal A.T.;
RT "Structure and regulation of the carB operon in Pseudomonas
RT aeruginosa and Pseudomonas stutzeri: no untranslated region exists."
RL J. Bacteriol. 176:2532-2542(1994).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; Pubmed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou P.,
RA Gaeber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.V., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an

RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch)
CC -----
DR EMBL: U04992; AAA19047.1; -
DR EMBL: U81259; AAB39251.1; -
DR EMBL: AE004889; AAG08143.1; -
DR InterPro: IPR001123; -
DR Pfam: PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 134 154 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT TRANSMEM 109 109 POTENTIAL.
FT CONFLICT 119 109 A -> G (IN REF. 1).
FT CONFLICT 119 119 P -> R (IN REF. 1).
FT CONFLICT 122 122 K -> Q (IN REF. 1).
SQ SEQUENCE 216 AA; 23248 MW; A30A08E71A5918BD CRC64;

Query Match 17.0%; Score 180.5; DB 1; Length 216;
Best Local Similarity 28.0%; Pred. No. 5.3e-10;
Matches 60; Conservative 41; Mismatches 90; Indels 23; Gaps 9;

QY 4 EWMFAVLTSTIITLSPGCAINTMTSLNHGYPAG-----GVYCMASDRFGDS--YCAG 56
DB 8 DFW-TYVIGVYFVILLRPNLSFLVLAISAQGVATGRACGV-----LGDVLMILS 60
QY 57 WRGVGTFSRSVIAFEVLKMGAAVILWLGIOQ---WRAAGAIDLKSLASTOSRRL-F 111
DB 61 ALGVAASLKAEPMLFVIGLGAALVFLYGVGMGRAMKRLNDEAFAQAQEVVDIOPF 120
QY 112 QRAVFNLTNKRISIVLEALFPQIFMP--QOPOLMOYIVLGVTIVDIIWIGVTTAQLA 169
DB 121 KRALILSNKRALIFETISFIQVDEYAVGL-SFIVLVILLEVLSALYSLIFITGV 179
QY 170 RIALWIKPKROMKA-LNKIFGSLFMVGLDASA 202
DB 180 RLAMFRRRQRLAAGATSGVGLPVGFKLATA 213

RESULT 5
YG27_SYNV3 STANDARD; PRT; 206 AA.
ID YG27_SYNV3
AC P74343; -
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 22.0 KDA PROTEIN SLR1627.
GN SLR1627.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=97061201; Pubmed=8905231;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hirosewa M., Sugitara M., Sasamoto S., Kimura T.,
RA Hosouchi T., Matsuno A., Muraki A., Nakazaki N., Naito K.,
RA Okumura S., Shimpō S., Takeuchi C., Wada T., Watanabe A.,
RA Yamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium

RA Nashimoto H., Saito N.;
 RL Submitted (SEP-1995) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [4]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RA Alba H., Baba T., Fujita K., Hayashi K., Honjo A., Horiuchi T.,
 RA Ikemoto K., Inada T., Isono K., Itoh T., Kanai K., Kasai H.,
 RA Kashimoto K., Kim S., Kimura S., Kitagawa M., Kitakawa M., Makino K.,
 RA Masuda S., Miki T., Mizobuchi K., Mori H., Motomura K., Nakamura Y.,
 RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sempel G., Seki Y.,
 RA Tagami H., Takemoto K., Wada C., Yamamoto Y., Yano M.;
 RL Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
 RN [5]
 RP IDENTIFICATION.
 RA Rudd K.E.;
 RL Unpublished observations (AUG-1994).
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: D13169; -; NOT_ANNOTATED_CDS.
 DR EMBL: D64044; -; NOT_ANNOTATED_CDS.
 DR EMBL: AE000344; AAC75631.1; -.
 DR EMBL: D90886; BA016464.1; -.
 DR ECoGene: EG12445; yf1K.
 KW Hypothetical protein; Transmembrane.
 FT TRANSMEM 8 28
 FT TRANSMEM 47 67 POTENTIAL.
 FT TRANSMEM 71 91 POTENTIAL.
 FT TRANSMEM 142 162 POTENTIAL.
 FT SEQUENCE 195 AA; 21248 MW; 5F86B828DDEDC090 CRC64;
 SO

Query Match 11.9%; Score 126; DB 1; Length 195;
 Best Local Similarity 24.5%; Pred. No. 6e-05;
 Matches 50; Conservative 36; Mismatches 80; Indels 38; Gaps 10;

QY 8 AYLLTSLITLTPSGSAINMTTSLNMGYPAGGVYCCASDRGDSVCACGR----- 58
 DB 8 AFWTYITLITANTPEPNNILASATSHGF-----RQSTRVLASGLFLIVMLI 56
 QY 59 GVGTLFSRSYI--AFEVLMWAGAAVLIWLGIQWRAGAIDKSLASTQSRHLLFORAV 115
 DB 57 CAGISFSLAVIDPAVAHLLSMAGAAIVWLIA--WKIATSPTEKD--GLQAKISWAS 111
 QY 116 FVNITNFKSVIFLAALFPQIPMQPOLMOYIVLGVTTIVVDIIVMIGVATLQRIALMT 175
 DB 112 ALQFVNKILLYGTALSTFVLPR-QTQALSMVY-GVSV---LLAMIG--TFGN--VCWA 161
 QY 176 KGPQKMAKLNKINGSLEMYGALL 199
 DB 162 LAGHLFQRLFPQYRGQNLIVLALL 185

RESULT 8
 RHTC_ECOLI

ID RHTC_ECOLI STANDARD: PRT: 206 AA.
 AC P27846;
 DT 01-AUG-1992 (Rel. 23, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUL-1999 (Rel. 38, Last annotation update)
 DE THREONINE EFFLUX PROTEIN.
 GN RHTC.
 OS *Escherichia coli*.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC *Escherichia*.
 OX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=92358234; PubMed=1379743;
 RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
 RT "Analysis of the *Escherichia coli* genome: DNA sequence of the region
 RT from 84.5 to 86.5 minutes.";
 RL Science 257:771-778(1992).
 RN [2]
 RP REVISIONS.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RA Mau B., Shao Y.;
 RT "The complete genome sequence of *Escherichia coli* K-12.";
 RL Science 277:1453-1474(1997).
 RN [3]
 RP SEQUENCE OF 1-107 FROM N.A.
 RC STRAIN-K12;
 RX MEDLINE=87115164; PubMed=3027506;
 RA Iriho N., Nakayama K., Nakayama H.;
 RT "The recQ gene of *Escherichia coli* K12: primary structure and
 RT evidence for SOS regulation.";
 RL Mol. Gen. Genet. 205:298-304(1986).
 RN [4]
 RP CONCEPTUAL TRANSLATION.
 RA Rudd K.E.;
 RL Unpublished observations (DEC-1997).
 RN [5]
 RP CHARACTERIZATION.
 RX MEDLINE=99313167; PubMed=10386596;
 RA Zakataeva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
 RA Livshits V.A.;
 RT "The novel transmembrane *Escherichia coli* proteins involved in the
 RT amino acid efflux.";
 RL FEBS Lett. 452:228-232(1999).
 CC -1- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE.
 CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
 CC -1- SIMILARITY: BELONGS TO THE RHT FAMILY.
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION. A FRAMESHIFT WAS
 CC INTRODUCED IN POSITION 80 TO PRODUCE THIS ORF.
 CC -1- CAUTION: REF.3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS
 CC IN POSITIONS 47 AND 73.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: M87049; AA67619.1; ALT_FRAME.
 DR EMBL: AE000456; AAC76826.1; ALT_FRAME.
 DR EMBL: M30198; -; NOT_ANNOTATED_CDS.
 DR PIR: S30713; S30713.
 DR ECoGene: EG11468; rhtC.
 DR InterPro: IPR001123; -.
 DR Pfam: PF01810; Lyse; 1.
 DR Transport; Transmembrane.
 KW

DR PROSITE: PS00218; AMINO-ACID-PEPMEASE_1; 1.
KW Transport; Amino-acid transport; Transmembrane; Inner membrane.
FT TRANSMEM 18 38 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 114 134 POTENTIAL.
FT TRANSMEM 135 155 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 278 298 POTENTIAL.
FT TRANSMEM 322 352 POTENTIAL.
FT TRANSMEM 356 400 POTENTIAL.
FT TRANSMEM 425 445 POTENTIAL.
SQ SEQUENCE 457 AA; 50215 MW; 5F2819526EA06BD8 CRC64;

Query Match 8.9%; Score 94; DB 1; Length 457;
Best Local Similarity 18.1%; Pred. No. 0.15;
Matches 50; Conservative 46; Mismatches 88; Indels 92; Gaps 10;

QY 3 LEWMPAYL-LTSIIILSPSGAINMTTSLNHYAGGYCWSADRTGDSYCAHMGV 61
DB 148 LEWMPFEEKVATIIIMVAGFII--IWGIGNGQPTGIHNLMSN---GGFFSNGW--LG 200
QY 62 TLFERSVIAFE-----VLKMGAGAYLIWLGIOQ 89
DB 201 MWMSLQWMPFAYGIEIGITAGEAKDPEKSIPRAINSVPMRLIVYVGLFVI-MSIYP 259
QY 90 WRAAGAILDKSLASTOSRRHLFORAV--FVNLTNPKSIV----- 126
DB 260 WNVGAGSPFVLTPOHMGITFAASILNFVLTASLSAINSDVFGVGRMLHGMAEGSAP 319
QY 127 -----FLAALFPOFIMPQOPOLMQYIVLGVTTIVDIYIMIGY 164
DB 320 KIFSRTSRGIPWTVLVMTTALLFAVLYINYPENVFVIAISLATFAFVWMIMILSQ 379
QY 165 ATLQRIALMIKPKOMKALN-KIFGSLFMLVGALL 199
DB 380 IAFRRRLP-----PEVKALKFKVPGGVATTIGLI 410

RESULT 13
PROY_SALTY STANDARD; PRT; 456 AA.
AC P37460;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PROLINE-SPECIFIC PERMEASE PROY.
GN PROY.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxID=602;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=LT2;
RX MEDLINE=97453475; PubMed=9308174;
RA Liao M.K., Gort S., Maloy S.;
RT "A cryptic proline permease in Salmonella typhimurium";
RL Microbiology 143:2903-2911(1997).

CC -1- FUNCTION: PERMEASE THAT IS INVOLVED IN THE TRANSPORT ACROSS THE
CYTOPLASMIC MEMBRANE OF PROLINE.
CC -1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN. INNER MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE AMINO ACID PERMEASE FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation-
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).

DR EMBL: X74420; CA52441.1; -
DR PIR: S35983; S35983.
DR StyGene; SG10314; PROY.
DR InterPro; IPR002027; -
DR Pfam; PF00324; aa-permeases; 1.
DR PROSITE: PS00218; AMINO-ACID-PEPMEASE_1; 1.
KW Transport; Amino-acid transport; Transmembrane; Inner membrane.
FT TRANSMEM 18 38 POTENTIAL.
FT TRANSMEM 39 59 POTENTIAL.
FT TRANSMEM 96 116 POTENTIAL.
FT TRANSMEM 117 137 POTENTIAL.
FT TRANSMEM 157 177 POTENTIAL.
FT TRANSMEM 198 218 POTENTIAL.
FT TRANSMEM 243 263 POTENTIAL.
FT TRANSMEM 278 298 POTENTIAL.
FT TRANSMEM 322 352 POTENTIAL.
FT TRANSMEM 356 376 POTENTIAL.
FT TRANSMEM 400 420 POTENTIAL.
FT TRANSMEM 425 445 POTENTIAL.
SQ SEQUENCE 456 AA; 50047 MW; EED5485A676F7783 CRC64;

Query Match 8.7%; Score 92; DB 1; Length 456;
Best Local Similarity 18.5%; Pred. No. 0.23;
Matches 51; Conservative 46; Mismatches 87; Indels 92; Gaps 11;

QY 3 LEWMPAYL-LTSIIILSPSGAINMTTSLNHYAGGYCWSADRTGDSYCAHMGV 61
DB 148 LEWMPFEEKVATIIIMVAGFII--VWGIGNGQPTGIHNLMSN---GGFFSNGW--LG 200
QY 62 TLFERSVIAFE-----VLKMGAGAYLIWLGIOQ 89
DB 201 MINSIQWMPFAYGIEIGITAGEAKDPEKSIPRAINSVPMRLIVYVGLFVI-MSIYP 259
QY 90 WRAAGAILDKSLASTOSRRHLFORAV--FVNLT-----NP 122
DB 260 WNVGAGSPFVLTPOHMGITFAASILNFVLTASLSAINSDVFGVGRMLHGMAEGSAP 319
QY 123 K-----SIVFLAALFPOFIMPQOPOLMQYIVLGVTTIVDIYIMIGY 164
DB 320 KVFAKTSRGIPIWTVLVMTTALLFAVLYINYPENVFVIAISLATFAFVWMIMILSQ 379
QY 165 ATLQRIALMIKPKOMKALN-KIFGSLFMLVGALL 199
DB 380 IAFRRRLP-----PEVKALKFKVPGGVATTIGLI 410

RESULT 14
LYSE_CORGL STANDARD; PRT; 236 AA.
AC P94633;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-DEC-1998 (Rel. 37, Last annotation update)
DE LYSINE EXPORTER PROTEIN.
GN LYSE.
OS Corynebacterium glutamicum (Brevibacterium flavum).
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacteriaceae; Corynebacteriaceae;
OC Corynebacterium.
OX NCBI_TaxID=1718;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=R127;
RX MEDLINE=97126810; PubMed=8971704;
RA Vrljic M., Sahm H., Eggeling L.;

RT "A new type of transporter with a new type of cellular function: L-
lysine export from Corynebacterium glutamicum";
RL Mol. Microbiol. 22:815-826(1996).
CC -1- FUNCTION: INVOLVED IN THE EFFLUX OF EXCESS OF L-LYSINE. THIS IS

QY	175	IKGPKOMALNKITFGSLIFMLVGALLASARIA	205
Db	116	IFVNLTNKSVIALAIFPQFMQOQLAOYLILGVTTVIVDMVWGTATLQRIANM	175
QY	175	IKGPKOMALNKITFGSLIFMLVGALLASARIA	205
Db	176	IKGPKOMALNKAFSGSLIFMLVGALLASARIA	206
RESULT	2		
Q9KVK7			
ID	Q9KVK7	PRELIMINARY:	PRT: 205 AA.
AC	Q9KVK7		
DT	01-OCT-2000	(Tremblrel. 15. Created)	
DT	01-OCT-2000	(Tremblrel. 15. Last sequence update)	
DT	01-OCT-2000	(Tremblrel. 15. Last annotation update)	
DE	HYPOTHETICAL PROTEIN VCO136.		
GN	VCO136.		
OS	Vibrio cholerae.		
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.		
OX	NCBI_TaxID=666:		
RP	SEQUENCE FROM N.A.		
RP	STRAIN=EL TOR N16961 / SEROTYPE O1;		
RX	MEDLINE=20406833; PubMed=10952301;		
RA	Heidelberg J.F., Eissen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,		
RA	Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,		
RA	Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,		
RA	Malnadzeva M.D., Vamatheyan J., Bass S., Qin H., Dragoi I., Sellers P.		
RA	McDonald L., Uterback T., Fleishmann R.D., Nieman W.C., White O.,		
RA	Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,		
RA	Fraser C.M.:		
RT	*DNA sequence of both chromosomes of the cholera pathogen Vibrio		
RT	cholerae.		
RL	Nature 406:477-483(2000).		
DR	EMBL: AE004104; AAF9313.1; -.		
DR	TIGR: VCO136.		
QO	SEQUENCE 205 AA: 22099 MW: AE42CA6FA6EA1223 CRC64:		

Query Match	43.3%;	Score 459.5;	DB 2;	Length 205;
Best Local Similarity	43.4%;	Pred. No. 3.3e-33;		
Matches 89;	Conservative 44;	Mismatches 69;	Indels 3;	Gaps 2

[illegible]

RC STRAIN-R1;
RX MEDLINE-20036896; PubMed-10567266;
RA White O., Eisen J.A., Heidelberg J. F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Hart D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Moffatt K.S., Qlt H., Jiang L., Pamphille W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Uppacher T., Zalewski C.,
RA Mekaroum K.A., Arvind L., Daly M.J., Milton K.W., Felschmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Paaser C.M.;
RT Genome Sequence of the Radiotolerant Bacterium Deinococcus
RT radiodurans R1.";
RL Science 286:1577-1577(1999).
RI EMBL: AF002037; AAF1548.1; -
DR TIGR: DR1989;
DR INTERPRO: IPR000719; -;
DR INTERPRO: IPR001125; -;
DR Pfam: PF01810; lysE; 1;
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
SQ SEQUENCE 241 AA; 25161 MW; A4017ABFEDB60374 CRC64;

Query Match	21.5%;	Score 228;	DB 2;	Length 241;
Best Local Similarity	30.3%;	Pred. No. 1.le-12;		
Matches 64;	Conservative 41;	Mismatches 84;	Indels 22;	Gaps 6;

```

OY      8  ATLLTLLTLLTSPGSAIINMTTSLNHGYPAGGYCMASDRTGSGYACGR-----G
OY      9  |: : : : |: : : : |: : : : |: : : : |: : : : |: : : : |: : : :
Db      38  AELVAAYVATLLPREGGLMYILLASLGG-----WAGISALGTGAGMYIVLASANG
OY      60  VGLTFRSVAIEVYLKAGAAIYILWIGCIQOMRAAIDLK---SLASTOSR--HLFORA
Db      91  LSLMIASSLASVYKTAGAYLITGLIKLISKALSSKEASLAAPAFOSRLITGG
OY      115  VEVNLNPKSIVTALAEPEQFIPEQPOFLM-QYIVSVTTIVDIIWIGVATLAAOTAL
Db      151  AMTELNPATLAFELIAPQGVNPAVGHVGGFLLGTTSVNTLAMLVATLAEIGCA
OY      174  WIKG-PKQMKALNKIPGSLPEWYGLLASGAR
Db      211  RIQGNPFGQGVKASGAMIALGTVAVER
RESULT      4
O9R6F8      PRELIMINARY;      PRT;      204 AA.
AC      O9R6F8;
DT      01-MAY-2000 (TREMblrel. 13, Created)
DT      01-MAY-2000 (TREMblrel. 13, last sequence update)
DT      01-JUN-2000 (TREMblrel. 14, last annotation update)
DE      TIORE101 PROTEIN.
GN      TIORE101.
OS      Agrobacterium tumefaciens (Ti plasmid pTi737).
OC      Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group.
OC      Rhizobiaceae; Agrobacterium.
OX      NCBI_TaxID=362;
RN      11]
RP      SEQUENCE FROM N.A.
RC      SPRAIN-WAFR301001;
RA      SUZUKI K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,
RT      Katoh A., Yoshida K.;
RL      Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.
RL      Gene 242:351-356(2000).
RN      12]
RP      SEQUENCE FROM N.A.
RC      SPRAIN-WAFR301001;
RA      SUZUKI K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.,
RT      "Novel structural difference between nopaline- and octopine-type
RT      gene:construction of genetic and physical map and sequencing of
RT      trb/tral and rep gene clusters of a new Ti plasmid pTi-SAKURA."
RL      Biochim. Biophys. Acta 1396:1-7(1998).
RN      13]
RP      SEQUENCE FROM N.A.

```

RC STRAIN=MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA(I): Strategy for DNA sequencing of a
RT Japanese cherry-Ti plasmid.";
RL Nucleic Acids Symp. Ser. 37:159-160(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (III): Characteristics of T-DNA.";
RL Nucleic Acids Symp. Ser. 39:185-186(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (IV): Characteristics of tra region.";
RL Nucleic Acids Symp. Ser. 39:187-188(1998).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (V): Complete nucleotide sequence of
RT plasmid pTi-SAKURA's vir region in Agrobacterium tumefaciens.";
RL Nucleic Acids Symp. Ser. 39:265-266(1998).
DR EMBL: AB016260; BAA87726.1; -
DR INTERPRO: IPR001123; -
DR Pfam: PF01810; LysE; 1.
KW plasmid
SQ SEQUENCE 204 AA; 21015 MW; A6E2AE267E0E97D0 CRC64;

Query Match 18.4%; Score 195.5; DB 2; Length 204;
Best Local Similarity 24.4%; Pred. No. 6.7e-10;
Matches 55; Conservative 44; Mismatches 83; Indels 43; Gaps 6;
QY 1 MTEWMAVLLTSLITLSPGSGAINMTTSLNHGYPAGGYCWASDRIGDSCAGMRCV 60
DB 1 MDISTLAFMAAFVFAFVAFASPGPNMTIVARTISHGAASGATV-----GA 44
QY 61 GTL-----FSSVIAFE-----VLKAGAAVLIWIGIQWRAGAI-DIKSLA 102
DB 45 GTVAGILLFTTLAFAFGSLTAEMGIWMTLLRYGAAYLIMWIKIMTAEPPVPELOPVS 104
QY 103 STOSRRLFORAVFNLTNPKSIVFLAALPQPI---MPQOQLMQYIVLGVTTIVDI 158
DB 105 ERRGLLAVFAIGVALNIGNKMPLEFVALLPNVGSSLLPCHTAIAAAYTL-----AVEM 159
QY 159 IVMIGVATLQAIRIALWIKGPKOMKALNKIFGSLFMLVGALLASAR 203
DB 160 VVIGGHVILAGRARAKLLRPTTVVRVNRRTAGVGAVVAVNASR 204
RESULT 5
Q9KFP9 PRELIMINARY; PRT; 207 AA.
AC Q9KFP9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE BH0429 (TREMBlrel. 15, Last annotation update)
GN BH0429. PROTEIN.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA Takami H., Nakasone K., Takaki Y.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP001508; BAB04148.1; -
SQ SEQUENCE 207 AA; 22317 MW; 9FED14C8E8DD3F0 CRC64;

Query Match 18.1%; Score 192.5; DB 2; Length 207;
Best Local Similarity 26.3%; Pred. No. 1.2e-09;
Matches 57; Conservative 50; Mismatches 87; Indels 23; Gaps 6;
QY 1 MTEWMAVLLTSLITLSPGSGAINMTTSLNHGYPAGGYCWASDRIGDSCAGMRCV 58
DB 1 MDISTLSTLFGAVALLTTPGPDILFVLAQSMQNRQAGIV-----TALGCTGLLVH 53
QY 59 -----GVGTFSSVIAFEVILKMGAAVLIWIGIQWRAG--AIDKSLASTQSRH 109
DB 54 ITAAVVGISALIVQSALAFVTVKYGAAVLLYLAKKAFQEKGEGLSIDKQ---TLAVGA 110
QY 110 LQGRVFNVLTPKSIIVFLAALFPQIPQOPQ-LMQYIVGVTTIVDIWIGVATLA 168
DB 111 IYKKGIINWLVNPKSLFLLALPQVNSGAGSAPWQMLLGLVFLIQAFITFSLVSWFA 170
QY 169 ORIA-LWIKGPKOMKALNKIFGSLFMLVGALLASARH 204
DB 171 EKVGGILMRSSFIKQMRIRKGGLLALIGLYVAFSKN 207

RESULT 6
Q9R6J5 PRELIMINARY; PRT; 229 AA.
AC Q9R6J5;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
DT 01-MAY-2000 (TREMBlrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE TIORF64 PROTEIN.
GN TIORF64.
OS Agrobacterium tumefaciens (Ti plasmid pTiT37).
OC Plasmid pTi-SAKURA.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Agrobacterium.
OX NCBI_TaxID=362;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Suzuki K., Hattori Y., Uraji M., Ohta N., Iwata K., Murata K.,
RA Katoh A., Yoshida K.;
RT "Complete nucleotide sequence of a plant tumor-inducing Ti plasmid.";
RL Gene 242:331-336(2000).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Suzuki K., Ohta N., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Novel structural difference between nopaline- and octopline- type trbu
RT gene: construction of genetic and physical map and sequencing of
RT trb/trai and rep gene clusters of a new Ti plasmid pTi-SAKURA.";
RL Biochim. Biophys. Acta 1396:1-7(1998).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Hattori Y., Suzuki K., Ohta N., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA(I): Strategy for DNA sequencing of a
RT Japanese cherry-Ti plasmid.";
RL Nucleic Acids Symp. Ser. 37:159-160(1998).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Ohta N., Suzuki K., Hattori Y., Uraji M., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (III): Characteristics of T-DNA.";
RL Nucleic Acids Symp. Ser. 39:185-186(1998).
RN [5]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;
RA Uraji M., Suzuki K., Ohta N., Hattori Y., Katoh A., Yoshida K.;
RT "Genome structure of pTi-SAKURA (IV): Characteristics of tra region.";
RL Nucleic Acids Symp. Ser. 39:187-188(1998).
RN [6]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF301001;

OC Bacteria; Proteobacteria; alpha subdivision; Sphingomonas group;
OC Zymomonas.
NCBI_TaxID=542;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ZM4;
RA Um H.W., Kang H.S.;
RT "The sequence analysis of 42D7 fosmid clone of Zymomonas mobilis
RT ZM4.";
RL Submitted (JUN-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF157493; AAD42407.1; -
DR INTERPRO; IPR001123; -
DR PFAM; PF01810; Lyse; 1.
KW Hypothetical protein.
SQ SEQUENCE 158 AA; 17454 MW; FE44A9639DA75A64 CRC64;

Query Match 17.3%; Score 184.5; DB 2; Length 158;
Best Local Similarity 30.1%; Pred. No. 4.7e-09;
Matches 46; Conservative 34; Mismatches 58; Indels 15; Gaps 5;

OY 59 GVGTLFERSVIAFEVLMKAGAAVILMLGIC-----QMRAGAIDLKSLASTOSRRHLFOR 113
DB 11 GLGAVLRASLAVNLKMSGAAVLCWGINLHPKKNLYDNDPSPMSATKALRGCF-- 68
OY 114 AVEFNLTFNPKSIYFLALPFOFIMPOOPOLMOYIVLGVTIYVDIIVMIGVATLAQRIA- 172
DB 69 --FTNINLPKVGIFVYFLPQF--MPSSHLIQYAF--LTFIHMTITLIFALLATAP 123
OY 173 --LWIKGPKOMKALNKIFGSLFMVAGALASAR 203
DB 124 LIRFLKRPRLSYIDRLGCVTAFGIKIALSK 156

RESULT 10
O9RMXO PRELIMINARY; PRT; 205 AA.
AC O9RMXO;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TREMBLrel. 14, Last annotation update)
DE PK02-63.
OS Bacillus anthracis.
OC Plasmid PK02.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RA Okinaka R.T., Cloud K., Hampton O., Hill K.K., Kelm P., Lamke G.,
RA Kumano S., Manter D., Martinez Y., Svensson R., Tatsum L.R.,
RA Brown A.E., Jackson P.J.;
RL EMBL; AF188935; AAF13668.1; -
DR EMBL; AF188935; AAF13668.1; -
DR INTERPRO; IPR001123; -
DR PFAM; PF01810; Lyse; 1.
KW Plasmid.
SQ SEQUENCE 205 AA; 22900 MW; 6DC91EB968B524B8 CRC64;

Query Match 17.3%; Score 184; DB 2; Length 205;
Best Local Similarity 29.5%; Pred. No. 7e-09;
Matches 64; Conservative 36; Mismatches 83; Indels 34; Gaps 10;

OY 1 MLEEMFAVILSIITLSPGSAINTMTTSLNHGYPAG-GVYCWASDRTPG----SYCA 55
DB 1 MDNINITVLVLTGTVISPGPMWAVILKNSL---YSRELGVSTVAGIATGSLIHIVCL 57
OY 56 GWRGVGLFERSVIAFEVLMKAGAAVILMLGICQWRRA---AAIDLKSLASTOSRRHLFO 112
DB 58 --IGIVITISKSLFNLTNLTWKIGAVALLTIGIKLRSKSPAAIKKNSTTWK--AFR 113
OY 113 RAVEVNLTPKRSIVFLALPFOFIMPOQPOLMOYIVLGVTIYVDI--VMIGVATLAQ 169

DB 114 SCFLTDMLNPKATFLYLAFTQVIEPNTNIFVQ-SVYGLTWSVBIHMHVLFELTHKS 172
OY 170 -----RIALMWIKGPKOMKALNKIFGSLFMVAGALIA 200
DB 173 VANNYFLSTHSWE-----RVGTALILIGIILA 200

RESULT 11
O9P920 PRELIMINARY; PRT; 213 AA.
AC O9P920;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBLrel. 15, Last annotation update)
DE AMINO ACID TRANSPORTER.
GN XF2730.
OS Xylella fastidiosa.
OC Bacteria; Proteobacteria; gamma subdivision; Xanthomonas group;
OC Xylella.
OX NCBI_TaxID=2371;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=9A5C;
RX MEDLINE=20365717; PubMed=10910347;
RA Simpson A.J.G., Rehnach F.C., Arruda P., Abreu F.A., Acencio M.,
RA Alvarenga R., Alves L.M.C., Araya J.E., Bala G.S., Baptista C.S.,
RA Barros M.H., Bonaccorsi E.D., Bordin S., Bove J.M., Briones M.R.S.,
RA Bueno M.R.P., Camargo A.C., Camargo L.E.A., Carraro D.M., Carter H.,
RA Coutinho L.B., Colombo C., Costa F.F., Costa M.C.R., Costa-Neto C.M.,
RA Coutinho L.L., Cristofani M., Dias-Neto E., Docena C., El-Dorri H.,
RA Facincani A.P., Ferreira A.J.S., Ferreira V.C.A., Ferro J.A.,
RA Friaga J.S., Franca S.C., Franco M.C., Frohne M., Furlan L.R.,
RA Garnier M., Goldman G.H., Goldman M.H.S., Gomes S.L., Gruber A.,
RA Ho P.L., Hohelsel J.D., Junqueira M.L., Kemper E.L., Kitajima J.P.,
RA Krieger J.E., Kuramae E.E., Laigret F., Lambais M.R., Leite L.C.C.,
RA Lemos E.G.M., Lemos M.V.F., Lopes S.A., Lopes C.R., Machado J.A.,
RA Machado M.A., Madeira A.M.B.N., Madeira H.M.F., Marinho C.L.,
RA Marques M.V., Martins E.A.L., Martins E.M.F., Matsukuma A.Y.,
RA Menck C.F.M., Miracca E.C., Miyaki C.Y., Monteiro-Vitorello C.B.,
RA Moon D.H., Nagai M.A., Nascimento A.L.T.O., Netto L.E.S.,
RA Nhani A., Jr., Nobrega F.G., Nunes L.R., Oliveira M.A.,
RA de Oliveira M.C., de Oliveira R.C., Palmieri D.A., Paris A.,
RA Peixoto B.R., Pereira G.A.G., Pereira H.A., Jr., Pesquero J.B.,
RA Quaggio R.B., Roberto P.G., Rodrigues V., de Rosa A.J.M.,
RA de Rosa V.E., Jr., de Sa R.G., Santelli R.V., Sawasaki H.E.,
RA da Silva A.C.R., da Silva A.M., da Silva F.R., Silva W.A., Jr.,
RA da Silveira J.F., Silvestri M.L.Z., Siqueira W.J., de Souza A.A.,
RA de Souza A.P., Terenzi M.F., Truffi D., Tsai S.M., Tsunako M.H.,
RA Vallada H., Van Slyks M.A., Verjovski-Almeida S., Vettore A.L.,
RA Zago M.A., Zatz M., Meidanis J., Secubal J.C.;
RT "The genome sequence of the plant pathogen Xylella fastidiosa.";
RL Nature 406:151-157(2000).
DR EMBL; AE004079; AAF85515.1; -
DR INTERPRO; IPR001123; -
DR PFAM; PF01810; Lyse; 1.
SQ SEQUENCE 213 AA; 23614 MW; E97242DD507C8065 CRC64;

Query Match 16.8%; Score 178.5; DB 2; Length 213;
Best Local Similarity 25.0%; Pred. No. 2.e-08;
Matches 51; Conservative 38; Mismatches 94; Indels 21; Gaps 4;

OY 9 YLNTSIITLSPGSAINTMTTSLNHGYPAG-----GVYCWASDRTPGSDYCAWGRG 59
DB 21 FILTIVMFSISPGPMWVILQOSKNGVTKGLAVLTGTEIGVFIVY-----ILTALG 72
OY 60 VGTLFERSVIAFEVLMKAGAAVILMLGICQWRRAAGAILDKSLASTOSRRHLFORAVFNL 119
DB 73 ISTVLKEPSSVITGLIGAGAVILYIAYLSWPRONASNQPTASRSSTYGTFTIGVGLNL 132
OY 120 TNPKSIVFLALPFOFIMPOQPOLMOYIVLGVTIYVDIIVMIGVATLAQRIALMWIKPK 179

AC 09XLA0: 09XLA0: (Tremblrel. 15, Created)
 DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
 DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
 DE HYPOTHETICAL PROTEIN VCA0846.
 GN VCA0846.
 OS Vibrio cholerae.
 OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
 OX NCBI_TaxID=666;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=EL TOR N16961 / SEROTYPE O1;
 RX MEDLINE=20406833; PubMed=10952301;
 RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,
 Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
 Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
 Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
 McDonald L., Utterback T., Fleischmann R.D., Niernan W.C., White O.,
 Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
 Fraser C.M.;
 RA *DNA sequence of both chromosomes of the cholera pathogen Vibrio
 RT cholerae.*;
 RL Nature 406:477-483(2000).
 DR EMBL: AE004412; AAF96744.1; -.
 DR TIGR: VCA0846; -.
 SQ SEQUENCE 204 AA; 21719 MM; D09887299659FBBB CRC64;

Query Match 13.5%; Score 143.5; DB 2; Length 204;
 Best Local Similarity 27.4%; Pred. No. 2.6e-05;
 Matches 46; Conservative 27; Mismatches 74; Indels 21; Gaps 5;

QY 1 MTEWMFAYLLNLSILTPSGAINTMTTSLNHGYPAGGVYCWASDRTGDSYCAG---- 56
 ||| | : :||| | | | : : |
 DB 1 MTLTWLSLFTICILGAMSPGSLAMVAKHSLAGRKNGFNAAMA-----HAFGIGYAF 55
 QY 57 --WRGVGTFSRSVIAFEVWKWGAAYLIWLGIOWRAGAIDLK-----SIATOSRRH 109
 | : : :| : : :||| | | | : : | : : |
 DB 56 ITLIGLAVYLHOSPYVFKTISTAGAAIYLLGMNALRKGAVAKLESSESYSVMQSAR- 114
 QY 110 LFGRAVFVNLTNPKSIVFLAALFQFIMPQOPOLMOYIVIGVTTIYVD 157
 : : :| : | : ||| | : : :| : |||
 DB 115 ---EGLLISLSPKIALFFIALFSQYAVGS-DLTSKAAIVITPLIYVD 158

Search completed: May 6, 2001, 14:40:50
 Job time: 246 sec

1

2

3

OM of: US-09-466-935-2 to: EST:* out_format : pfs
Date: May 6, 2001 3:09 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 Compugen Ltd.

Command line parameters:

-MODBL-frame p2n model -DPY-XLP
-O-/g9n2.1/USFO.spool/US09466935/runat_03052001.075709.14787/app.query.fasta.1.529
-DB-EST -QEXT=fastap -SUFFIX=est -GAPOP=12.000 -GAPEXT=4.000
-MINMATCH=0.100 -LOOPL=0.000 -LOOPEXT=0.000 -OGAPOP=4.500
-OGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -DELOP=6.000
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELDP=6.000
-DELEXT=7.000 -START=1 -MATRIX=blonum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR SCORE=pct -THR MAX=100 -THR_MIN=0
-ALIGN=15 -MODE=LOCAL -OUTP=pfis -NORM=ext -MINLEN=0
-MAXLEN=2000000000 -USER=US09466935 -CGCN1_1.3199 -NCPU=6
-ICPU=3 -LONGLOG -NO_XLPHY -WAIT -THREADS=1

Search information block:

Query: US-09-466-935-2

Query length: 205

Database: EST*

Database sequences: 9623517

Database length: 73081774

Search time (sec): 1674.270000

score list:

Sequence	Strd Orig	Zscore	EScore	len	! Documentation
gb_gss12:AO859250	-	227.00	455.61	28e-16	473 ! AO859250 nbe001114r CUGI Rice
gb_est19:BG038404	+	180.50	359.83	6.0e-11	443 ! BG038404 dg34h08.y1 Xenopus lae
gb_est19:BF614897	+	164.50	323.01	4.0e-09	428 ! BF614897 dg18a02.y1 Xenopus lae
gb_gss12:AO365906	+	149.50	292.80	3.2e-07	562 ! AO365906 nbx0004N05f CUGI Rice
gb_gss12:AO858095	+	126.00	240.49	0.0003	769 ! AO858095 nbe001114f CUGI Rice
gb_gss12:AZ576789	-	111.50	217.56	0.0050	407 ! AZ576789 01a12 Shot-gun genomic
gb_est18:BF784988	+	95.50	175.02	1.117	932 ! BF784988 601478429P1 NIH_MGC_6
gb_est18:BF526220	+	90.50	166.73	3.40	774 ! BF526220 602071232P1 NCI_CGAP_E
gb_est13:BW960930	+	89.00	166.36	3.57	607 ! BW960930 EST992083 GVN Medicago
gb_est14:BF202999	+	87.00	165.88	3.79	438 ! BF202999 WHE1768.B10_C2025 Whea
gb_gss14:AZ601442	-	86.00	162.57	5.80	489 ! AZ601442 IM0419P08R Mouse 10kb
gb_gss14:AO240216	-	86.00	162.57	6.18	511 ! AO240216 CIT-HSP-2385P1.TF.1 CT
gb_est12:BF043998	+	86.00	162.01	6.24	514 ! BF043998 BP25004A20H3 Soares r
gb_est18:AL509238	+	86.00	159.84	8.23	623 ! AL509238 AL509238 Hordeum vulg
gb_est19:BF679835	+	86.00	154.77	15.77	977 ! BF679835 602154419P1 NIH_MGC_83
gb_gss14:AO95402	+	85.50	152.16	22.03	1123 ! AO95402 AF095402 Salmonella t
gb_est14:AW290151	+	85.00	158.98	9.19	559 ! AW290151 NXN012607F Nsf Xylem
gb_est16:BG128201	+	84.50	158.19	10.17	547 ! BG128201 EST473847 tomato shoot
gb_est16:BG128201	+	84.50	146.22	47.21	1583 ! BG128201 602350810P1 NIH_MGC_9
gb_est12:BE334024	-	83.50	156.75	12.23	517 ! BE334024 us29d09.y1 Soares NMEH
gb_gss14:AO991044	-	83.50	153.88	17.68	667 ! AO991044 Rfc01899 Photobabds
gb_est10:AW701541	-	83.50	153.07	19.62	717 ! AW701541 ug87d10.y1 NCI_CGAP_Lu
gb_est11:AI515059	+	83.00	154.24	16.87	589 ! AI515059 LD46737.Sprime LD Dros
gb_est18:AI256881	+	83.00	152.88	20.10	665 ! AI256881 LP03567.Sprime LP Dros
gb_est14:AA941801	+	83.00	152.44	21.25	702 ! AA941801 AD26506.Sprime LD Dros
gb_est18:BF493242	+	83.00	152.27	21.74	702 ! BF493242 AT01557.Sprime AT Dros
gb_est18:BF489020	+	83.00	150.37	25.60	786 ! BF489020 AT24710.Sprime AR Dros
gb_est17:BG112141	-	83.00	150.37	27.74	831 ! BG112141 602333261P1 NCI_CGAP_M
gb_gss17:BE525982	-	82.50	153.13	19.46	593 ! BE525982 HS_5329.A2 A10_77A RBC
gb_est17:BE731315	-	82.50	148.07	34.48	881 ! BE731315 601566433P1 NIH_MGC_21
gb_est18:BF128485	-	82.00	154.42	13.41	2265 ! BF128485 601810323P1 NIH_MGC_4
gb_est19:AV651857	+	82.00	154.29	16.28	478 ! AV651857 AV651857 GIC Homo sapi
gb_est19:AV651904	+	82.00	154.22	16.78	488 ! AV651904 AV651904 GIC Homo sapi
gb_est1:AA156246	+	82.00	153.44	18.70	536 ! AA156246 z150e12.s1 Soares preg
gb_est19:AA156246	+	82.00	153.32	24.54	625 ! AA156246 z150e12.x1 NCI_CGAP_P
gb_est19:BE049512	+	82.00	149.43	31.28	751 ! BE049512 xW88B12.N1 NCI_CGAP_P
gb_est11:AA735580	+	82.00	149.43	47.37	1001 ! AA735580 LD21566.Sprime LD Dros
gb_est16:BF345786	-	81.50	149.19	30.63	675 ! BF345786 602017856P1 NCI_CGAP_L
gb_gss12:AO083242	-	81.50	149.51	30.63	675 ! AO083242 RBC111-54M8.TK RBC1-11
gb_gss12:AO641276	+	81.50	149.51	44.85	879 ! AO641276 RBC193-DpnII-29E21.7J
gb_est18:BE745367	+	81.50	146.62	40.85	879 ! BE745367 601578883P1 NIH_MGC_9
gb_est10:BF683697	-	81.50	145.00	55.21	1015 ! BF683697 602139957P1 NIH_MGC_4

seq_name: gb_gss12:AO859250

seq_name	seq	documentat	block	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
gb_gss12:AO859250	-	81.00	153.30	19.04	443 ! BE439347 CDO1015 WHE05E08R I					
gb_est19:BF612680	+	80.50	148.00	37.59	647 ! BF612680 daa20a09.y1 NICH D X					
gb_est12:AL503330	-	80.50	147.11	42.12	700 ! AL503330 AL503330 Hordeum vu					
gb_est17:BE412566	-	80.50	144.76	56.90	862 ! BE412566 MCG002.F1IR990625 I					

seq documentat

seq	documentat	block	LOCUS	DEFINITION	ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM
AO859250	473 bp	DNA	nbe001114r	CUGI Rice BAC library (ECORI) Oryza sativa genomic clone nbe001114r, DNA sequence.	AO859250	1	GI:6209707	GSS.	Oryza sativa
AO859250	1	GI:6209707	GSS.	Oryza sativa	Oryza sativa	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta; Magnoliophyta; Liliopsida; Poales; Poaceae; Ehrhartoideae; Oryzae; Oryza.			

REFERENCE
Wing, R.A. and Dean, R.A.
A BAC End Sequencing Framework to Sequence the Rice Genome
Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seg primer: GGAAACGCTATGACCATG
Class: BAC ends
High quality sequence start: 41
High quality sequence stop: 415.
Location/qualifiers
1. 473
/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="nbe001114r"
/clone_lib="CUGI Rice BAC library (ECORI)"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACindigo; Site 1: EcorI; Site 2: EcorI; Rice is the most important food crop in the world. Half of the world population, especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety using EcorI as the cloning enzyme. The library contains 55,296 clones with an average insert size of 121 Kb providing approximately 15 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Three high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening and can be requested from the Clemson University BAC/EST Resource Center (www.genome.clemson.edu)."

alignment_scores: 227.00 Length: 53
Ratio: 4.633 Gaps: 1
Percent Similarity: 92.453 Percent Identity: 86.792

alignment_block:
US-09-466-935-2 x AQ859250/rev ..

Align seg 1/1 to reverse of: AQ859250 from: 1 to: 473

```
1 MetThrLeuGluTrpTrpPheAlaTyrLeuLeuThrSerIleIleLeuTh 17
|||||
229 ATGACCTTAGATGTTGTTGCTTACCCTGCTGACATGACATCTTTAAG 180
17 rLeuSerProGlySerGlyAlaIleAsnThrMetThrThrSerLeuAsnH 34
|||||
179 CCTGTGACAGAGCTCTGCTGCATATCAACACATATACACCTCGCTCAAC 130
34 ISGLTYTPROAGLGLY..GLYVALTYCYSTTPALASERASPRALTYHrg 50
|||||
129 ACGGTATACGGCGCGCGATTAAGTCTATGCTGGGCTTCAGACCGGACTG 80
50 IYASPSER 52
|||||
79 GCGATTCG 72
```

seq_name: gb_est95:BG038404

seq_documentation_block: 443 bp mRNA EST 24-JAN-2001
LOCUS BG038404
DEFINITION dg34h08.y1 Xenopus laevis gastrula non normalized Xenopus laevis
CNA clone XENOPUS.SOURCE.ID:xlnnga009p16 5' similar to TR:Q9RSX2
Q9RSX2 CONSERVED HYPOTHETICAL PROTEIN. ; mRNA sequence.

ACCESSION BG038404
VERSION BG038404.1 GI:12480989

KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis

REFERENCE 1 (bases 1 to 443)
AUTHORS Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Page,D.,
Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
,B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
Washu Xenopus EST project, 1999
Unpublished (1999)

TITLE JOURNAL
COMMENT Other ESTs: dg34h08.x1
Contact: Sandy Clifton, Ph.D
Washu Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu

Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: <http://www.resgen.com/> Please reference the id listed
below when ordering this clone: Source lab clone id - xlnnga009p16
Seq primer: -40RP from Gldoo
High quality sequence stop: 430.
Location/Qualifiers

FEATURES
Source

1..443
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XENOPUS.SOURCE.ID:xlnnga009p16"
/clone_id="Xenopus laevis gastrula non normalized"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F"
/note="Vector: pBluescript SK-; Site.1: EcoRI; Site.2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA (equal
parts from stage 10.5 and stage 11.5 gastrulae).
EcoRI-XhoI cut cDNA was then ligated into Unizap-XR

(Stratagene) with EcoRI at the 5' end and XhoI at the 3' end. SS-library phagemids were prepared by mass excision from the original library and normalized by hybridization to biotinylated driver (prepared from the same library by PCR) to Cot-omega of 11. After removal of hybrids and excess driver by streptavidin sepharose chromatography, the ss-phagemids were made double stranded and electroporated into Top-10 F'. Original library construction by Bruce Blumberg (Cho et al. 1991 Cell 67, 1111-1120). Note: This is a Xenopus Gene Collection (XGC) library."

BASE COUNT 74 a 142 c 127 g 100 t
ORIGIN

alignment_scores:
Quality: 180.50 Length: 141
Ratio: 1.920 Gaps: 2
Percent Similarity: 66.667 Percent Identity: 28.369

alignment_block:
US-09-466-935-2 x BG038404 ..

Align seg 1/1 to: BG038404 from: 1 to: 443

```
9 TyrLeuLeuThrSerIleIleLeuThrLeuSerProGlySerGlyAlaI 25
:::|||||:|||||:|||||:|||||:|||||:|||||:|||||:
14 TTCCATACATCATCTTATGTTGCCCGCTCCGGCAGCGGAGATG 63
25 easnThrMetThrThrSerLeuAsnHISGLTYTPROAGLGLYVALT 42
:::|||||:|||||:|||||:|||||:|||||:|||||:
64 TTACACGCTGCTGCCGCTTTCAGGAGGCGCGGAGACGTCATC 113
42 YCYSTTPALASERASPRALTYHGLYASPSERTYCYs..... 54
|||||
114 CGGCTTTCGgc.....TGCAGCTGGGCAATC 139
55 .....AlaGlyTPRArgLYaLYaLYaThrLeuPheSerAr 66
|||||:|||||:|||||:|||||:|||||:
140 GTGCCGATCTGCTCGCGCCATCACCGGCTTTCGCGCATCTCCACAC 189
66 gserValIleAlaPheGluValLeuLeuThrPAlaGlyAlaIaTyrLeu 83
|||||:|||||:|||||:|||||:|||||:|||||:
190 CAGTGGCGCTTTCAGGCTGCTCAATATCTGGGCGTCCCTATCTGC 239
83 IeTPLeuGlyIleGlnGlnTrpArgAlaAlaGlyAlaIleAspleuLys 99
|||||:|||||:|||||:|||||:|||||:
240 TCTACATGGCGTGAACACGCTGACAGAGATGGCGCTGAAGATCGAT 289
100 SerLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValP 116
:::|||||:|||||:|||||:|||||:|||||:
290 GAACGAGGCGACCGCGAAGACCGCGCGTATCGTGAAGATCTCT 339
116 eValAsnLeuThrAsnProLysSerIleValPheLeuAlaIlePheP 133
|||||:|||||:|||||:|||||:|||||:
340 CATCAATATGCTGTAACCGCAACTGTCATCTTTTGGCTTCCTGC 389
133 roGlnPheIleMetProGlnGln 140
|||||:|||||:|||||:|||||:
390 CGCAGTTCATCGCGCGCATGAG 412
```

seq_name: gb_est89:BF614897

seq_documentation_block: 428 bp mRNA EST 14-DEC-2000
LOCUS BF614897
DEFINITION dg18a02.y1 Xenopus laevis gastrula non normalized Xenopus laevis
cDNA clone XENOPUS.SOURCE.ID:xlnnga005b04 5' similar to TR:Q9RSX2
Q9RSX2 CONSERVED HYPOTHETICAL PROTEIN. ; mRNA sequence.

ACCESSION BF614897
VERSION BF614897.1 GI:11787968

KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis

genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from *Oryza sativa*, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation a particular sequence with a probability of 99.9 %. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT 91 a 151 c 180 g 140 t
ORIGIN

alignment_scores:
Quality: 149.50 Length: 108
Ratio: 2.534 Gaps: 2
Percent Similarity: 54.630 Percent Identity: 27.778

alignment_block:
US-09-466-935-2 x AQ365906 ..

Align seg 1/1 to: AQ365906 from: 1 to: 562

```

1 MethrLeuGluTrpTPheAlaTyrlLeuThrSerlleleLeuTh 17
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
299 ATGTCGCGTGAACGTCGCGTTCGCGCTGTTGCGCTGTTGATCAG 348
17 rleuSerProGlySerGlyAlaIleAsnThrMetThrThrSerleuAsnH 34
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
349 TCTTCTCCGGCGCCAGCGCATTCGTCGATGTCAGTGCAGTCTGCAGT 398
34 lsglyTyPrAlaGlyGlyValTyrcyStrPaIaseraPaGlyThGly 50
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
399 ACGGCTTC.....TGGCGCGGTACTGMAACGCCCTCGGTCGCA 406
51 AspSerTyrcysAlaGlyTyPrAg..... 58
407 .....TGGCGCGGTACTGMAACGCCCTCGGTCGCA 438
59 .....GlyValGlyThrL 63
439 ATTGGCGCTGATCGTTCAGATCGCATCAGCCGCGAGCGCTGGTCCG 488
63 eupheSerArgSerValIleAlaPheGlyValLeuLysTrpAlaGlyAla 79
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
489 TGCTGCGCCAGCTCGCGCGCTGCGTTCACGTCGATCAATGTTGCGGCT 538
80 AlaTyrlLeuIleTrpLeuGlyIle 87
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
539 GCTTACCTGGTGTATCTGGGTGTG 562

```

seq_name: gb_gss12:A0858095

seq_documentation_block:

LOCUS A0858095 769 bp DNA GSS 03-NOV-1999
DEFINITION nbcd001114f CGI Rice BAC Library (EcoRI) *Oryza sativa* genomic
clone nbcd001114f, DNA sequence.

ACCESSION A0858095
VERSION A0858095.1 GI:6208552
KEYWORDS GSS.
SOURCE *Oryza sativa*.
ORGANISM *Oryza sativa*

REFERENCE 1 (bases 1 to 769)
AUTHORS Wing,R.A. and Dean,R.A.
TITLE A BAC End Sequencing Framework to Sequence the Rice Genome
JOURNAL Unpublished (1998)
COMMENT Contact: Wing RA
Clemson University Genomics Institute

Clemson University
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7288
Fax: 864 656 4293
Email: rwing@clemson.edu
Seq primer: TAATACGACTCACTATAGGCG
Class: BAC ends
High quality sequence start: 42
High quality sequence stop: 454.
Location/Qualifiers

FEATURES
source

```

1..769
/organism="Oryza sativa"
/strain="Japanica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone_lib="nbcd001114f"
/clone_1lb="CGI Rice BAC Library (EcoRI)"
/tissue_type="leaf"
/lab_host="E. coli DH10B"
/note="Vector: pBACindigo, Site_1: EcoRI, Site_2: EcoRI.
Rice is the most important food crop in the world. Half of
the world population, especially those inhabiting highly
populated areas of the humid tropics and subtropics, rely
on rice as their primary source of carbohydrate.
Monocotyledonous rice is a diploid plant (2n=24) with a
haploid genome equivalent of 431 Mbp (Arumuganathan and
Earle, 1991). The relatively small genome of rice, three
times larger than that of Arabidopsis, makes it suitable
for genomic studies. In order to facilitate positional
cloning, physical mapping and genome sequencing of rice,
we have constructed a BAC library from Oryza sativa. The
Nipponbare variety using EcoRI as the cloning enzyme. The
library contains 55,296 clones with an average insert size
of 121 Kb providing approximately 15 haploid genome
equivalents. The deep coverage allows the isolation a
particular sequence with a probability of 99.9 %. Three
high density filters, each containing 18,432 clones
(doubly spotted), represent the whole library for colony
screening and can be requested from the Clemson University
BAC/EST Resource Center (www.genome.clemson.edu)."
```

BASE COUNT 203 a 216 c 176 g 169 t 5 others
ORIGIN
alignment_scores:
Quality: 126.00 Length: 36
Ratio: 3.500 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 91.667

alignment_block:
US-09-466-935-2 x A0858095 ..

Align seg 1/1 to: A0858095 from: 1 to: 769

```

1 MethrLeuGluTrpTPheAlaTyrlLeuThrSerlleleLeuTh 17
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
658 ATGACCTTAGATGTGGTTCGCTTACCTGTCGATGATCATTTAAG 705
17 rleuSerProGlySerGlyAlaIleAsnThrMetThrThrSerleuAsnH 34
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
706 CCTGTGCCAGGCTCTGTGTCATCAACACATATGACCACTGCTCACC 755
34 lsglyTyPr 36
|||||  |||||  |||||  |||||  |||||  |||||  |||||  |||||
756 ACGGGTAT 763

```

seq_name: gb_gss23:A2576789

seq_documentation_block:

LOCUS A2576789 407 bp DNA GSS 08-DEC-2000
DEFINITION 01a12 Shot-gun genomic library of *Rhizobium* strain ANU265
sp. NGR234 genomic clone 01a12, DNA sequence.
ACCESSION A2576789


```

VERSION      AZ576789.1  GI:11603030
KEYWORDS     GSS
SOURCE       Rhizobium sp. NGR234.
ORGANISM     Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
              Rhizobiaceae; Rhizobium.
REFERENCE    1 (bases 1 to 407)
AUTHORS      Viprey, V., Rosenthal, A., Broughton, W.J., and Perret, X.
TITLE        Genetic snapshots of the Rhizobium species NGR234 genome
JOURNAL      Genome Biology, vol 1 (6), 0014.1-0014.7 (2000)
COMMENT      Contact: Virginie Viprey
              Laboratoire de Biologie Moleculaire des Plantes Superieures
              University of Geneva
              1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
              Tel: +44(0)1603450000
              Fax: +44(0)1603450045
              Email: virginie.viprey@bbserc.ac.uk
              Class: shotgun.
FEATURES     Location/Qualifiers
              1..407
               /organism="Rhizobium sp. NGR234"
               /strain="ANU265"
               /db_xref="taxon:394"
               /clone="01a12"
               /clone_lib="Shot-gun genomic library of Rhizobium strain
              ANU265"
              85 a 126 c 137 g 55 t 4 others
               /note="vector: M13; derivative strain of NGR234 cured of
              pNGR234a"
BASE COUNT   85 a 126 c 137 g 55 t 4 others
ORIGIN
alignment_scores:
  Quality: 111.50      Length: 135
  Ratio: 1.394        Gaps: 4
  Percent Similarity: 59.259 Percent Identity: 28.148
alignment_block:
US-09-466-935-2 x AZ576789/rev ..
Align seg 1/1 to reverse of: AZ576789 from: 1 to: 407
5 TrpTrpPheAlaTyrLeuLeuThrSerIleLeuLeuThrLeuSerProG1 21
||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
395 TGCGTGTCTTCACCTTCCTTCGCGCATCATCATCGTC.....CCGG 352
21 ySerGlyAlaIleAsnThrMetThrThrSerLeuAsnHisGlyTyrProA 38
| : : : : : : : : : : : : : : : : : : : : : : : : : : :
351 CATGGACATGGTCTTCGTCGCCAACGCGATACCGCGCGACGCGT 302
38 laGly.....GlyValTyrCysTrpAlaSerAspArgThr 49
: ||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
301 CCGCGCTCCGCGGCGGCGCATCATCGAGCGCGGCTGCTGCACACG 252
50 GlyAspSerTyrCysAlaGlyTrpArgGlyValGlyThrLeuPheSerAr 66
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
251 .....CTCATGCGCG.....CTCGCGCTCAGCGTGTCTTCATCT 214
66 gSerValIleAlaPheGluValLeuLysTrpAlaGlyAlaAlaTyrLeuI 83
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
213 GGTGCCGAGCTCTTCAAGTGTCTGCTCGCGCGCGCGCTACATCG 164
83 lefTrpLeuGlyrIleGlnInTrpArgAlaAlaGlyAlaIleAspLeuLys 99
||| : ||| : ||| : ||| : ||| : ||| : ||| : ||| : |||
163 CCTGGATCGGCTTCTCGCTGCTGCGAGTTCATCATCATTCATTCGCGCGCTC 114
100 SerLeuAlaSerThrGlnSerArgArgHisLeuPheGlnArgAlaValph 116
: : : : : : : : : : : : : : : : : : : : : : : : : : :
113 GAGCGGGGCGAGGCTCTCGCGCTGGGCGAGCTTTCGACAGGGGCGCT 64
116 eValAsnLeuThrAsnProLysSerIleValPheLeuAlaAlaLeuPheP 133
: : : : : : : : : : : : : : : : : : : : : : : : : : :
63 GACGAGCCTGATGAANCCGAGGCTATCNCNTTCATGTTCCGCGCNCATCC 14

```

```

133 roGln 134
|||||
13 CGCAG 9
seq_name: gb_est78:BE784988
seq_documentation_block:
LOCUS      BE784988      932 bp      mRNA      EST      20-OCT-2000
DEFINITION 601478429f1 NIH_MGC_68 Homo sapiens cDNA clone IMAGE:3881191 5',
            mRNA sequence.
ACCESSION  BE784988
VERSION    BE784988.1  GI:10206186
KEYWORDS   EST.
SOURCE     human.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 932)
AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL   Unpublished (1999)
COMMENT    Contact: Robert Strausberg, Ph.D.
            Tel: (301) 496-1550
            Email: Robert_Strausberg@nih.gov
            Tissue Procurement: DCTD/DTP/Gazdar
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: L1AM9649 row: j column: 08
            High quality sequence stop: 602.
FEATURES   Location/Qualifiers
            1..932
             /organism="Homo sapiens"
             /db_xref="taxon:9606"
             /clone="IMAGE:3881191"
             /clone_lib="NIH_MGC_68"
             /tissue_type="large cell carcinoma"
             /lab_host="DH10B (phage-resistant)"
             /note="Organ: lung; Vector: pCMV-SPORT6; Site: 1: NotI;
            Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
            Average insert size 1.8 kb. Library constructed by Life
            Technologies."
BASE COUNT 256 a 192 c 259 g 225 t
ORIGIN
alignment_scores:
  Quality: 95.50      Length: 233
  Ratio: 0.974       Gaps: 14
  Percent Similarity: 42.060 Percent Identity: 24.034
alignment_block:
US-09-466-935-2 x BE784988 ..
Align seg 1/1 to: BE784988 from: 1 to: 932
1 MetThrLeuGluTrpTrp.....PheAlaTyrLe 10
: : : : : ||||| : : : : :
193 CTCACACTCCCATGGTGGAGGACAGACAGATGATGATGATGATGATGA 242
10 uLeuThrSerIleIleLeuThrLeuSerProGlySerGlyAlaIleAsnT 27
||| : : : : ||| : : :
243 CCTCTACTGTGTGGCTTAAAGAATG..... 267
27 hrMetThrThrSerLeuAsnHisGlyTyrProAlaGlyGlyValTyrCys 43
: : : : : : : : : : : : : : : : : : : : :
268 .....TGGAACTACTCTTTCCGGTGGTGATCAATACAGC 303
44 TrpAlaSerAspArgThrGlyAspSerTyrCysAlaGlyTrpArgGlyva 60

```



```

seq_name: gb_est53:AW980930
seq_documentation_block: 607 bp mRNA EST 07-SEP-2000
LOCUS AW980930
DEFINITION EST392083 GVN Medicago truncatula cDNA clone pgvn-60i23, mRNA
sequence.
ACCESSION AW980930
VERSION AW980930.1 GI:8172475
KEYWORDS EST.
SOURCE barrel medic.
ORGANISM Medicago truncatula
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids I;
Fabales; Fabaceae; Papilionoideae; Medicago.
REFERENCE 1 (bases 1 to 607)
AUTHORS Fedorova,M., Pierson,B.L., Samac,D.A., Vance,C.P., Gantt,J.S., Peng
,H., Ellis,L., Town,C.D., Bowman,C.L., Craven,M.B., Hansen,T.S.,
Holt,I.E. and Fraser,C.M.
ESTs from one month old nitrogen-fixing root nodules of Medicago
truncatula
JOURNAL Unpublished (2000)
COMMENT Contact: Carroll P. Vance
Department of Agronomy and Plant Genetics
University of Minnesota
411 Borlaug Hall, 1991 Upper Buford Circle, St. Paul, MN 55108 USA
Tel: 612-625-7219
Fax: 651-649-5058
Email: vance004@maroon.tc.umn.edu
TIGR sequence name: M256764e
University of Minnesota name: M256764e
More information is available at: http://chrysie.tamu.edu/medicago
Seq primer: SKmod (CTA GAA CTA gta gat CC).
FEATURES
source
location/Qualifiers
1..607
/organism="Medicago truncatula"
/cultivar="genotype Al7"
/db_xref="taxon:3880"
/clone="pgvn-60i23"
/clone_lib="GVN"
/tissue_type="N2-fixing root nodules"
/dev_stages="effective root nodules harvested one month
post inoculation with Sinorhizobium meliloti"
/lab_host="E. coli strain XLOLR"
/note="Vector: phuescript SK-; Site_1: ECORI; Site_2:
XhoI; cDNA was prepared from polyA+ enriched RNA from
effective root nodules harvested one month post
inoculation with Sinorhizobium meliloti. The cDNA was
directionally ligated into the Uni-ZAP XR vector from
Stratagene and packaged using Gigapack III Gold packaging
extracts. Plasmids containing cDNA inserts were excised
from the recombinant lambda-ZAP phage using Ex-Assist
helper phage and propagated in XLOLR cells."
BASE COUNT 172 a 161 c 133 g 141 t
ORIGIN
alignment_scores:
Quality: 89.00 Length: 202
Ratio: 0.918 Gaps: 12
Percent Similarity: 48.020 Percent Identity: 22.772
alignment_block:
US-09-466-935-2 x AW980930/rev ..
Align seg 1/1 to reverse of: AW980930 from: 1 to: 607
48 ArgThrGlyAspSerTyrCysAlaGly.....Tr 57
||| |||:||||:|||||
569 AGAGGGGTCCCTCATGTTGGGGGAGGATCCACCACCTGCTTTG 520
|||:||||:||||:||||:|||||
57 pArgGlyValGlyThrLeuPheSerArgSerValIleAlaPheGluValL 74
||||:||||:||||:||||:|||||
519 GAGGAGCATATCCACCTGATTGGAGGGGCGTATCCGCC.....T 479

```

```

74 euLysTrpAlaGlyAlaAlaTyrLeuIleTrpLeuGlyIleGln.... 89
||| |||:||||:||||| |||:||||:|||||
478 TGATTGGAGGGCGTATCCAGCTGTGATTGGAGGGAGTATCCAGATG 429
|||:||||:||||:||||:||||:||||:||||:||||:||||:
90 ...TrpArgAlaAlaGlyAlaIleAspLeuLys.....SerLe 101
|||:||||:||||:||||:||||:||||:||||:||||:||||:
428 ATTGGAGGGGAGTATCCAGCTGTGATTGGTAGGAGGTATCCGCCAGCAT 379
||||:||||:||||:||||:||||:||||:||||:||||:||||:
101 uAlaSerThrGlnSerArg..... 108
||||:||||:||||:||||:||||:||||:||||:||||:||||:
378 TGTGGGAGGGTATCCAGAAGAATTGTTGGGAGGTATCCAGAAGATG 329
||||:||||:||||:||||:||||:||||:||||:||||:||||:
109 .....HisLeuPhe 111
||||:||||:||||:||||:||||:||||:||||:||||:||||:
328 TTGGAGGTGGACCAAGATTCTGCACCTGGAGGCCCTACTCTGCATGCTCT 279
|||:||||:||||:||||:||||:||||:||||:||||:||||:
112 GlnArgAlaValPhe.....ValAsnLe 119
|||:||||:||||:||||:||||:||||:||||:||||:||||:
278 GTTCACCCACATTTTCCTTCTCTTTCCTTTCCTTTCCTTTCCTTTCCT 229
|||:||||:||||:||||:||||:||||:||||:||||:||||:
119 uThrAsnProLysSerIleValPheLeuAlaAlaLeuPheProGlnPheI 136
|||:||||:||||:||||:||||:||||:||||:||||:||||:
228 TAGAGGCCT.....ATCATTTGGCGCTGTTCTTCATTAGCAGCAGCA 185
|||:||||:||||:||||:||||:||||:||||:||||:||||:
136 leMetProGlnGlnProGlnLeuMetGlnTyrIleValLeuGlyValThr 152
|||:||||:||||:||||:||||:||||:||||:||||:||||:
184 TTGTTCTCT...AACCCACCTCTCATG.....ATACTGGGATCATA 147
||||:||||:||||:||||:||||:||||:||||:||||:||||:
153 ThrIleValValAspIle.....IleValMetIleG1 163
|||:||||:||||:||||:||||:||||:||||:||||:||||:
146 TGAACCTGCTTGGCCATTAATGAAGGGCTCACCTCCATAAATTTTCCTCC 97
|||:||||:||||:||||:||||:||||:||||:||||:||||:
163 yTyAlaThrLeuAlaGlnArgIleAlaLeuTrpIleLysGlyProLysG 180
|||:||||:||||:||||:||||:||||:||||:||||:||||:
96 TCACATTCAAAATGAAGATCAGG.....AAGAACCCTCGA 62
|||:||||:||||:||||:||||:||||:||||:||||:||||:
180 lMetLysAlaLeuAsnLysIlePheGlySerLeuPheMetLeuValGly 196
|||:||||:||||:||||:||||:||||:||||:||||:||||:
61 ACCTT.....GGCAGTTCTTTCAGTTGTAGGA 33
|||:||||:||||:||||:||||:||||:||||:||||:||||:
197 AlaLeu 198
|||
32 AAGCTC 27
seq_name: gb_est84:BF202999
seq_documentation_block: 438 bp mRNA EST 06-NOV-2000
LOCUS BF202999
DEFINITION WHE1768_B10_C20S Wheat pre-anthesis spike cDNA library Triticum
aestivum cDNA clone WHE1768_B10_C20, mRNA sequence.
ACCESSION BF202999
VERSION BF202999.1 GI:11117741
KEYWORDS EST.
SOURCE bread wheat.
ORGANISM Triticum aestivum
Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
Magnoliophyta; Liliopsida; Poales; Poaceae; Poideae; Triticeae;
Triticum.
REFERENCE 1 (bases 1 to 438)
AUTHORS Anderson,O.D., Chao,S., Choi,D.W., Close,T.J., Fenton,R.D., Han
,P.S., Hsia,C.C., Kang,Y., Lazo,G.R., Miller,R., Rausch,C.J.,
Seaton,C.L. and Tong,J.C.
The structure and function of the expressed portion of the wheat
genomes - Pre-anthesis spike cDNA library
JOURNAL Unpublished (2000)
COMMENT Contact: Olin Anderson
US Department of Agriculture, Agriculture Research Service, Pacific
West Area, Western Regional Research Center
800 Buchanan Street, Albany, CA 94710, USA
Tel: 5105959773
Fax: 5105959818
Email: oanderson@pw.usda.gov

```


us-09-466-935-2.rst

Mon May 7 11:08:09 2001

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 6, 2001, 15:09:39 ; Search time 1239.41 Seconds
(without alignments)
7353.642 Million cell updates/sec

Title: US-09-466-935-3_COPY_187_804
Perfect score: 618
Sequence: 1 atgttgatgtattttcttcac.....ttcatttgattatttcggg 618

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1283235 seqs, 7373929652 residues
Total number of hits satisfying chosen parameters: 2566470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

- GenEmbl:*
- 1: gb_ba1.*
 - 2: gb_ba2.*
 - 3: gb_ba3.*
 - 4: gb_in1.*
 - 5: gb_in2.*
 - 6: gb_in3.*
 - 7: gb_om.*
 - 8: gb_ov.*
 - 9: gb_pat1.*
 - 10: gb_pat2.*
 - 11: gb_ph.*
 - 12: gb_pl1.*
 - 13: gb_pl2.*
 - 14: gb_pl3.*
 - 15: gb_pl4.*
 - 16: em_ba1.*
 - 17: em_ba2.*
 - 18: em_fun.*
 - 19: em_htgo_hum.*
 - 20: em_htgo_inv.*
 - 21: em_htgo_rod.*
 - 22: em_htg_hum1.*
 - 23: em_htg_hum2.*
 - 24: em_htg_hum3.*
 - 25: em_htg_hum4.*
 - 26: em_htg_hum5.*
 - 27: em_htg_hum6.*
 - 28: em_htg_hum7.*
 - 29: em_htg_hum8.*
 - 30: em_htg_inv1.*
 - 31: em_htg_inv2.*
 - 32: em_htg_other.*
 - 33: em_htg_rod.*
 - 34: em_hum1.*
 - 35: em_hum2.*
 - 36: em_hum3.*
 - 37: em_hum4.*
 - 38: em_hum5.*
 - 39: em_hum6.*
 - 40: em_hum7.*
 - 41: em_in.*
 - 42: em_om.*
 - 43: em_or.*

- 44: em_ov.*
- 45: em_pat.*
- 46: em_ph.*
- 47: em_pl.*
- 48: em_ro.*
- 49: em_sts.*
- 50: em_sy.*
- 51: em_un.*
- 52: em_vi.*
- 53: gb_sts1.*
- 54: gb_sts2.*
- 55: gb_sts3.*
- 56: gb_sy.*
- 57: gb_un.*
- 58: gb_vil.*
- 59: gb_vil2.*
- 60: gb_htg1.*
- 61: gb_htg2.*
- 62: gb_htg3.*
- 63: gb_htg4.*
- 64: gb_htg5.*
- 65: gb_htg6.*
- 66: gb_htg7.*
- 67: gb_htg8.*
- 68: gb_htg9.*
- 69: gb_htg10.*
- 70: gb_htg11.*
- 71: gb_htg12.*
- 72: gb_htg13.*
- 73: gb_htg14.*
- 74: gb_htg15.*
- 75: gb_htg16.*
- 76: gb_htg17.*
- 77: gb_htg18.*
- 78: gb_htg19.*
- 79: gb_htg20.*
- 80: gb_htg21.*
- 81: gb_htg22.*
- 82: gb_htg23.*
- 83: gb_htg24.*
- 84: gb_htg25.*
- 85: gb_pr1.*
- 86: gb_pr2.*
- 87: gb_pr3.*
- 88: gb_pr4.*
- 89: gb_pr5.*
- 90: gb_pr6.*
- 91: gb_pr7.*
- 92: gb_pr8.*
- 93: gb_pr9.*
- 94: gb_ro1.*
- 95: gb_ro2.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	618	100.0	840	9 AX030177	AX030177 Sequence
2	606	98.1	11509	1 AE000458	AE000458 Escherich
3	594	96.1	91414	2 EC00W85	M87049 E. coli gen
4	592.4	95.9	10592	1 AE005613	AE005613 Escherich
5	557.2	90.2	279589	75 AC073768	AC073768 Mus muscu
6	486	78.6	212936	66 AC020970	AC020970 Mus muscu
7	442.6	71.6	96086	3 STYSTMD1	AF233324 Salmonell
8	415.8	67.3	268294	66 AC020885	AC020885 Mus muscu
9	339	54.9	290452	76 AC079167	AC079167 Mus muscu
10	295.8	47.9	2695	2 ECORECQ	M30198 E.coli recQ
11	242.8	39.3	62274	66 AC020833	AC020833 Mus muscu

TITLE
JOURNAL
Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE
AUTHORS
3 (bases 1 to 11509)
TITLE
JOURNAL
Direct Submission
Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
608-263-7459
REFERENCE
AUTHORS
4 (bases 1 to 11509)
TITLE
JOURNAL
Direct Submission
Submitted (13-OCT-1998) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
On Sep 9, 1997 this sequence version replaced gi:1790254.
This sequence was determined by the E. coli Genome Project at the
University of Wisconsin-Madison (Frederick R. Blattner, director).
Supported by NIH grants HG00301 and HG01428 (from the Human Genome
Project and NCHGR). The entire sequence was independently
determined from E. coli K12 strain MG1655. Predicted open reading
frames were determined using GeneMark software, kindly supplied by
Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA,
30332 [e-mail: mark@ember.gatech.edu]. Open reading frames that
have been correlated with genetic loci are being annotated with CG
Site Nos., unique ID nos. for the genes in the E. coli Genetic
Stock Center (CGSC) database at Yale University, kindly supplied by
Mary Berlyn. A public version of the database is accessible
(http://cgsc.biology.yale.edu). Annotation of the genome is an
ongoing task whose goal is to make the genome sequence more useful
by correlating it with other data. Comments to the authors are
appreciated. Updated information will be available at the E. coli
Genome Project's World Wide Web site
(http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
its annotations are periodically updated; this is version M34. No
sequence changes. Annotation updates: updated gene identifications
and products; all new functional assignments courtesy of Monica
Riley; added promoters, protein binding sites, and repeated
sequences described in reference 1. The unique numeric identifiers
beginning with a lowercase 'b' assigned to each gene (protein- or
RNA-encoding) are now designated as gene synonyms instead of
labels. This should allow them to be searched for in Entrez as gene
names.

FEATURES
source
1. .11509
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
81. .950
/gene="pldA"
/note="b3821"
81. .950
/gene="pldA"
/EC_number="3.1.1.32"
/function="enzyme; Central intermediary metabolism:
Phosphorus compounds"
/note="o289; 99 pct identical to PA1_ECOLI SW: P00631;
CGSCNo. 384"
/codon_start=1
/transl_table=11
/product="outer membrane phospholipase A"
/protein_id="AAC76824.1"
/db_xref="GI:2367300"
/translations="MTIQGMLLPVEMLPWVAQATVKEVHDPAVRGSIANNLQ
EHNPTFLPYDNYLIYQTSDLNKEIASTDWAENARKDEKVEQLSLAPFLWRGIL
GPNVSITQKSWQLNSNSPFRETNYEPQLFLGFDYDFAGWTLRDVDMGY
NHDNSGRSDPTSRWNLRYTLRLMAENGNLVEKPYVVGNDNDPDTIKYMGYYQLK
IGYHLGDVILSAKGYNWNTGYGGAELGSLPYITKVRLYTVQVSGESLIDYNFNG
TRYGVGVMLNDLF"
965. .993
promoter

promoter
1019. .1047
/note="factor Sigma70; predicted +1 start at 4003392"
/note="factor Sigma70; predicted +1 start at 4003446"
1077. .2909
/gene="recQ"
/note="b3822"
1077. .2909
/gene="recQ"
/EC_number="3.6.1.1"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/note="o610; 99 pct identical to 607 amino acids
of RECQ_ECOLI SW: PI5043 but has 3 additional N-ter
residues"
/codon_start=1
/transl_table=11
/product="Atp-Dependent DNA helicase"
/protein_id="AAC76825.1"
/db_xref="GI:2367301"
/translation="NNVAQAEVLNLESGAKOVLOFTFGYQQFRPQOEI IDTVLSGRD
CLVMPETGGKSLCYQIPALLANGLTVVVSPILSLMKDQVDQLOANGVAAACLNSTOT
RECOLBVTGCGTQIRLLYIAPERLMDLNFLEHLAHPNVPVLLAVDEAHCLISONGHDF
LMEKFKPLDLMRYVQEQKSGI IYCNRAKVEDTAARLQSKLSIAAAYHAGLNW
RADVQKRFQDDLOIVVATVAFGMIKPNVRFVHFDPINIESYQTCRGRGDL
PAMAMLYFPADMAWLRLEKPGQGLQIERHKLNAMGAFABEQTCRRLVLLNRYFG
EGHDEPCGNCIDCLDPKQYDQALSTIGRVNORFGMGYVVEVIRGANNQRT
DYGHDKLVYGMGRKSHHWYSVIROLIHGLVNTONTAQHSALQLEAARPVAESS
LQLAVPRIVALPKAKQKSGFNDRKLFKLRKLSIADESNNVPPVVFENDATLIE
MAQMPITASEMLSVNGVGMKRLERFPFMAIIRAHVVDGDEE"
2104. .2140
/note="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
3162. .3189
/note="factor Sigma70; predicted +1 start at 4005588"
3224. .3592
/gene="yigJ"
/note="b3823"
3224. .3592
/gene="yigJ"
/function="orf; Unknown"
/note="o322; sequence change shortens and
changes N-terminus relative to earlier version (YIGJ_ECOLI
SW: P27846)"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76826.1"
/db_xref="GI:2367302"
/translation="MGYQMLGALKKEAVSAPQVELAKSGRSFLKGLLNLNPKA
IIFYGVSFLFVGDNVGTTRWGIIFALIIVETLAWFTVVASLFPALPQMRRGYQRLAKW
IDGFAGALFAGFGIHLIISR"
3612. .3645
/note="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
complement(3654. .4070)
/gene="yigK"
/note="b3824"
complement(3654. .4070)
/gene="yigK"
/function="orf; Unknown"
/note="f138; 100 pct identical to YIGK_ECOLI SW: P27847"
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76827.1"
/db_xref="GI:1790258"
/translation="MIAFVILKWAAGAVLIWLGIQWRAAGAILDKSLASTQSRRLP
QRAVFNLINPKSIVLAAALFPQFIMPQOQLMQIYIVGVTTIVVDIIVMIGIYATLQ
RIALWTKGPKQKALNKIFGSLFMLVGLALLASARHA"
4200. .4230
/note="factor Sigma70; predicted +1 start at 4006629"
complement(4227. .4254)
repeat_region
promoter
gene
CDS
repeat_region
gene
CDS
promoter
promoter

/note="factor Sigma70; predicted +1 start at 4006612"
4384..5406
/gene="pldB"
/note="b3825"
4384..5406
/gene="pldB"
/EC_number="3.1.1.5"
/function="enzyme; Central intermediary metabolism:
Phosphorus compounds"
/note="o340; 99 pct identical amino acid sequence and
equal length to PLDB_ECOLI SW: P07000"
/codon_start=1
/translation="lysophospholipase L(2)"
/protein_id="AAC76828.1"
/db_xref="GI:2367303"
/translation="MF00K0WETRENAFAAFTMGPLTDFWRQDEAEFTGVDDIPVR
FVFRRAQHDVVICPGRIESVYKVAELAYDLHGFVDVLIIDHROGRSGVLLADP
HLGVNRFNDYDDLAAPWQEQVQPGPKRYILAHSMGSAISTLFQRHPGVCDIA
LTAPMFGIVIRPFSFARQILNWAEPFRDGYAIGTGRWALPFAINLVHRSORY
RRLNRFYADPTIRVGGPTVHWRESILAGEQVLAGAGDDATPTLLQAEERVQNR
MDRFFCELRTAAGHPVEGGRPLVIKAYHEILFEKDAVASVALHAIYVDFNHRNSPSG
NRSTEV"
5381..5408
/note="factor Sigma70; predicted +1 start at 4007807"
5426..5435
/note="factor Sigma70; predicted +1 start at 4007854"
5438..5466
/note="factor Sigma70; predicted +1 start at 4007865"
5494..5523
/note="factor Sigma70; predicted +1 start at 4007922"
5526..6041
/gene="yigL"
Query Match 98.1%; Score 606; DB 1; Length 11509;
Best Local Similarity 99.8%; Pred. No. 3.9e-167;
Matches 617; Conservative 0; Mismatches 0; Indels 1; Gaps 1;
Qy 1 atgttgatgtattcttcacgcgtgcacattggtgcacattggtggcgcttatgagcccggt 60
Db 2973 ATGTTGATGTTATTCTCACCGTCGCCATGTGTGCACATGTGGCGCTTATGAGCCCGGT 3032
Qy 61 cccgattctttttgtctctcacgcgtgtcagtcgtctccgttaaaagacgatgatg 120
Db 3033 CCCGATTTCTTTTGTCTCTCACACCCCTGTCTAGTCGTCCTCCGTAAGACGATGATG 3092
Qy 121 ggcgtgctggcgattacacgcgcgtgaatggtttggcggtggattgcgctggtgcctg 180
Db 3093 GCGGTGCTGGGCATTTACCTGCGCGGTAATGTTTGGGCTGGGATTCGCTGCTGGCCTG 3152
Qy 181 catttgattatcgaaaaatggcctggctgcatacgtgattgattggtggcggtgacctg 240
Db 3153 CATTGTGATTCGAAAAAATGGCTGGCTGCATACGCTGATTTATGGTGGCGGTGG-CTG 3211
Qy 241 tatctctgtagtggtttaccagatgctacgtggtgacactgaaagagagcggtttct 300
Db 3212 TATCTCTGCTGGATGGGTACCAGATGCTACGTGTGCACACAAAAGAGCGGTTCCT 3271
Qy 301 gcacctgcccacaggtgcagctggcgaaaagtgggcgagtttccctgaaaggttactg 360
Db 3272 GCACCTGCGCCACAGTGCAGTGGCGAAAAAGTGGCGCAGTTTCTCTGAAAGGTTTACTG 3331
Qy 361 accaatctcgtaatccgaaagcattatctacttgcctcggtgtcttcattgtttgctc 420
Db 3332 ACCAATCTCGGTAAATCCGAAGCGGATTAFTACTTTGGCTCGGTGTTCTCATGTTTGTCTC 3391
Qy 421 ggtgataacgttggcactacgcgcgcgtggggcatttttggcgctgacattgctcgaagc 480
Db 3392 GGTGATACGTTGGGCATACCGCGCGCTGGGCAATTTTGGCTGCTGATCATTTCTCGAAACG 3451
Qy 481 ctggcggtggtttaccgctcgcttgcagcgcgtgttttgcctgcgcgaatgcgcgcgtggttat 540
Db 3452 CTGGCGGTGTTTACCGTGTGTTGCCAGCCTGTTTGGCCCTGCCCAATGCCGCGGTGTTAT

Qy 541 caacgtctgcgaagtggattgatgttttgcggcggttatttgcggatttggcatt 600
Db 3512 CAACGCTTGGCGAAGTGGAATTGATGTTTTCGGGGCGGTATTTCGCCGATTTGGCATT 3571
Qy 601 catttgattatttcgcgg 618
Db 3572 CATTGATTATTTCGCGG 3589
RESULT 3
ECOUW85 91414 bp DNA BCT 29-MAY-1995
LOCUS E. coli genomic sequence of the region from 84.5 to 86.5 minutes.
DEFINITION M87049
ACCESSION M87049.1 GI:836656
VERSION M87049.1
KEYWORDS
SOURCE
ORGANISM Escherichia coli.
Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia
1 (bases 1 to 91414)
REFERENCE
AUTHORS Daniels,D.L., Plunkett,G. III., Burland,V. and Blattner,F.R.
TITLE Analysis of the Escherichia coli genome: DNA sequence of the region from 84.5 to 86.5 minutes
JOURNAL Science 257 (5071), 771-778 (1992)
MEDLINE 92358234
REFERENCE 2 (bases 1 to 91414)
AUTHORS Plunkett,G. III., Burland,V., Daniels,D.L. and Blattner,F.R.
TITLE Analysis of the Escherichia coli genome. III. DNA sequence of the region from 87.2 to 89.2 minutes
JOURNAL Nucleic Acids Res. 21 (15), 3391-3398 (1993)
MEDLINE 93347969
REFERENCE 3 (bases 1 to 91414)
AUTHORS Blattner,F.R., Burland,V., Plunkett,G. III., Sofia,H.J. and Daniels,D.L.
TITLE Analysis of the Escherichia coli genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes
JOURNAL Nucleic Acids Res. 21 (23), 5408-5417 (1993)
MEDLINE 94089392
REFERENCE 4 (bases 1 to 91414)
AUTHORS Rudd,K.E., Sofia,H.J., Koonin,E.V., Plunkett,G. 3rd, Lazar,S. and Rouviere,P.E.
TITLE A new family of peptidyl-prolyl isomerases
JOURNAL Trends Biochem. Sci. 20 (1), 12-14 (1995)
MEDLINE 95184296
REFERENCE 5 (bases 1 to 91414)
AUTHORS Daniels,D.L.
TITLE Direct Submission
JOURNAL Submitted (12-FEB-1992) Donna L. Daniels, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
Email: ecoligenetics.wisc.edu; Phone: 608-262-2534; Fax: 608-263-7459
COMMENT On May 29, 1995 this sequence version replaced gi:148169.
This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. The entire sequence was independently determined from E. coli MG1665. Overlaps and conflicts with other sequence determinations are annotated. Reference [1] describes the original sequence determination of a 91408 bp sequence. References [2], [3], and [4] describe subsequent corrections and/or updates to that sequence.
FEATURES
source
1. 91414
/organism="Escherichia coli"
/strain="K-12"
/sub_strain="MG1655"
/db_xref="taxon:562"
/clone.lib="lambda library of Daniels and Blattner"
/map="bp 1 at 3974 kb; 84.7 minutes"
/note="This sequence comprises the following lambda clones: DD850(EC15-116), DD851(EC13RM4), DD854(EC15-177), DD856(EC19RM8.1), DD858(EC17RM9),

DD861(EC17-221),DD864(EC27-149), DD865(EC21-76),
DD867(EC27-860). M13mp19 or Janus vectors were used for
subcloning."

1..200
/note="corresponds to K02846 (1..200)"
1..5863
/note="clone EC15-116"
complement(1..4899)
/note="corresponds to M37337 (1..4900)"
1..339
/note="corresponds to V00349 (164..500)"
10..86
/gene="aspt"
10..86
/note="CGSC No. 989"
/gene="aspt"
95..170
/gene="trpt"
/note="CGSC No. 66"
95..170
/gene="trpt"
177..205
/note="terminator-like sequence"
253..396
/note="predicted bend of 81 degrees"
complement(266..862)
/gene="fl98"
complement(266..862)
/gene="fl98"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AAA67566.1"
/db_xref="GI:148170"
/translation="WAHTSRHNEFSIGASALWECMLNOWLGRLYQNDQHTGLQFEA
RIAQKOSLVQIHERQDLILLITAPKPMDEFSSQLIGYFTLALTSAPSKLGDLNLY
RLEWPDFQOHEAGLEADIVPILTTSSAELAQOQIAMLNGCTWLPVSWARKGSLHT
VVDSTLSRPLAIWLQNSDKNALIRDLKINVLDEV"
586..724
/note="predicted bend of 77 degrees"
complement(703..1104)
/gene="fl33"
complement(703..1104)
/gene="fl33"
/note="CGSC No. 18010; possible pssr gene"
/codon_start=1
/transl_table=11
/protein_id="AAA67567.1"
/db_xref="GI:148171"
/translation="MDELLKTFLEVSRTSRHFGRAESLYLTOSAVSFRIQLENQLG
VNLFTRHNRNRLTAAGKILLPYAETLMTSQAAKRWIRPHDITSLFSLVPAPRCGNV
CLISGDACIKIRMPIQAYSKRRLPNSLW"
complement(1136..1165)
1151..1179
/note="promoter matrix score of 63"
/note="promoter matrix score of 57"
1223..1636
/gene="ol37"
1223..1636
/gene="ol37"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AAA67568.1"
/db_xref="GI:148172"
/translation="MAESFTTNNRYFDKNKHYPGRGFSRHGDDTKEAOLLERHGYAFNE
LDLGRKPEVTEELKFAVACRGREPVTEAERVMSKYMTRIKRKRHTLSAVNRLLK
VLKTPILTNKKGMPFYYVSSFCRWMSKRSIAR"
1562..1586
/gene="ol37"
/note="terminator-like sequence"
complement(1585..3135)
/gene="f516"

misc_feature
misc_feature
misc_feature
misc_feature
tRNA
gene
tRNA
gene
terminator
misc_structure
gene
CDS
misc_structure
gene
CDS
promoter
promoter
gene
CDS
terminator
gene

CDS
complement(1585..3135)
/gene="f516"
/note="similar to Mg chelatase subunit of Rhodobacter
capsulata"
/codon_start=1
/transl_table=11
/function="unknown"
/protein_id="AAA67569.1"
/db_xref="GI:148173"
/translation="MHIMVTQGLMSLSIVHTRAALGVNAPPIITVEVHISKGLPLTM
VGLPETTVKEARDVRSAIINSGYEYPAKTIINLAPADLPKEGGRVDLPITAIALLA
SEOLTANKLDSEYELVGEALTGALGVPCAISSEATEAKSGKKIIVAKDNEDEVLIN
GEGCLIAHLQAVCAFLLEGKHALEXPPTDSAEALQHDLSDVIGQGGKRGLEITAA
GHNILLIGPPTGKTMLASINGLLIPDLSNEEALSAAILSVNAESVQKWRQRF
RSRPHASLTAMVGGAIPGPGEISLAHNGVLFDELPEFERRTLDALRPTIESGQTH
LSRTRAKITYPARQFVAAMNPSPTGHYQGNHNRCTPEOTLRYLNRLSGPFLDPRDLS
LEIPLPPGILSKTVVPGESSATVKQVMAAREQKRONKLNALDLSPEIRQPKLE
SEDAMWLEGTLIHGLSIRAWQRLKLVARTIADIQSDITTRQHLQEAIVSTRAIDRL
IHLQKLLT"
3287..3433
/note="predicted bend of 79 degrees"
3297..3326
/note="promoter matrix score of 74"
3336..3676
/note="corresponds to M11651 (1..341)"
3336..3737
/note="corresponds to V00289 (1..402)"
3336..12791
/note="corresponds to M10313 (1..9456)"
3379..5721
/note="V00290; ECILVX(1..2343)"
3458..3556
/gene="ilvL"
3458..3556
/gene="ilvL"
/codon_start=1
/transl_table=11
/product="ilvGEDA operon leader peptide"
/protein_id="AAA67570.1"
/db_xref="GI:148174"
/translation="MTALLRVISLVISVVVVIIPCGAALGRKA"
3696..4679
/gene="ilvG"
3696..4679
/gene="ilvG"
/note="CGSC No. 603; inactive in wild-type E. coli K12"
/codon_start=1
/transl_table=11
/product="acetoxyhydroxy acid synthase II, large subunit"
/protein_id="AAA67571.1"
/db_xref="GI:148175"
/translation="MNGAOWVHALRAQGVNTVFGVGGAIMPVYDALYDGGVHELLC
RHEQGAAMAIGYARATGTCVCIATSGCATNLITGLADALLDSIPVVAITQVSAP
FIGTDAFQEVYDVLGSLACHTSFVQSLEELPRMAEAFDVACSRGPGVLDIPKD
IQLASGLPEWFTTIVENAVFPAEVEQAQAKPMLYVGGVGMAQVAPLRE
FLAATKMPATCTILGLGAEADYPYILGMLGHMGTAAANFAVQECDLLIAVGAFDDR
VTGLNTFAPHASVIHMDIDPAEMNKLRQAHVALQGLNALLPALQOPLNQ"
4675..5340
/gene="o221"
4675..5340
/gene="o221"
/note="the mutation ilv02096 joins this ORF to ilvG,
restoring function missing in wild-type"
/pseudo
/codon_start=1
replace(4677..4678,"tatg")
variation
Query Match 96.1%; Score 594; DB 2; Length 91414;
Best Local Similarity 99.7%; Pred. No. 1.5e-163;
Matches 616; Conservative 0; Mismatches 0; Indels 2; Gaps 2;
QY 1 atgttgatgtattctccacgtgcacattggtcacattggtgcgttatgaccccggt 60
|||||

Db	60863	ATGTTGATGTTATTTCTCACCGTCGCCATCGTGTGCACATTTGTGGCGCTTATGAGCCCGGT	60922	FEATURES	Location/Qualifiers
Qy	61	ccqgattctctttttgtctctcagacgcctgtcagtcagtggtcccgtaagaagcgatgatg	120	source	1..10592
Db	60923	CCCGATTTCTTTTGTCTCTCAGACCGCTGTCTAGTCGTCTCCCGTAAAGAAGCGATGATG	60982		/organism="Escherichia coli O157:H7"
Qy	121	ggcgtgctggcattacactcgcggcgtaatagggtttggcgctgggattgcgctgttggccctg	180	misc_feature	/strain="EDL933"
Db	60983	GGCGTGTGGGCATTTACCTGGCGGCTAATGGTTTGGCGTGGGATGGCGTCTTTGGCGCTG	61042		/serotype="O157:H7"
Qy	181	catttgattatcgaaaaagcctggcctgcatacgcctgattatggcgcggtggccctg	240	gene	/db_xref="taxon:83334"
Db	61043	CATTTGATATTCGAAAAATGGCCCTGGCTGCTACGCTGATATATGGTGGCGGTGG-CTG	61101	CDS	/note="enterohemorrhagic"
Qy	241	tatctctgctggatgggtttaccagatgctacgtggtgcactgaaataagagcggtttct	300		<1..71
Db	61102	TATCTCTGCTGGATGGGTTACACAGATGCTACGTGTGTGCACATGAAAAAGAGCGGTTTCT	61161		/note="O-island #155; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
Qy	301	gcacctgcgcacaggctcgagctcgcgaaataagtgggcgagcttctcgaaggtttactg	360		complement(72..836)
Db	61162	GCACCTGCGGCACAGGTGAGCTGGCGAAAGATGGGCGACGTTTCCCTGAAA-GTTTACTG	61220		/gene="Z5332"
Qy	361	accaatctcgtaatccgaaagcagatctatctactttgctcggtgttctcattgtttgtc	420		/gene="Z5332"
Db	61221	ACCAATCTCGTATATCCGAAAGCGATATCTACTTTGCTCGGTGTTCTCATTTGTTGTC	61280		/function="orf; Unknown function"
Qy	421	ggtgataacgttggcactaccgcgcgtggggcatttttgcgtgcatcatttgcgtgatactgcgaaacg	480	gene	/notes="Residues 1 to 131 of 254 are 95.41 pct identical to residues 1 to 131 of 161 from Escherichia coli K-12 Strain MG1655; B3815"
Db	61281	GGTGATAACGTTGGCAGTACCGCGCGCTGGCGCATTTTGGCGCTGATCATTTGCGAAACG	61340	CDS	/codon_start=1
Qy	481	ctggcgtggtttacagctggttgcagcctgtttccctgcgcgaataatgcgcgtggttat	540		/transl_table=11
Db	61341	CTGGCGTGGTTTACCGTGGTTGCGAGCTGTTTGCCCGTCCGCAATGCGCGTGGTTAT	61400		/protein_id="AAG59008.1"
Qy	541	caacgtctggcgaagtgcattgatgtgtttccggggcggttatttgcggatttggcatt	600		/db_xref="GI:12518686"
Db	61401	CAACGCTGCGCAAGTGGATGATGTTTTCGCGGGCGGTTATTTGCGGGATTTGGCATT	61460		/translation="MLSAFQLENRLRLEVEESQPLVNAVWIDLVEPDDDERLRVOS ELGQSLATRPELEDIEASARFFEDDDGLHIHFFFEADAEDHAGNSVAFVTRDGRLLF TRERELPAFLRYRMARSQSMVDGNAYELLLDFETKIEQLADEIENIYDLBQLSR VIMGHOGDEYDEALSTLAELEDIGWKVRLCLMDTQALNLFVRKARLPGQLQARE ILRDIESILPHNESLFOKVNFLMQAAMGFINEONRIKIFSVSVVFLPPTLVASSY GMNFEFPELKLWISFGYGAIFMILAGLAPLYLFRKNWL"
Qy	601	catttgattatttcgcgg	618	misc_feature	2157..4848
Db	61461	CATTTGATTTATTCGCG	61478	gene	/note="O-island #156; Region of the EDL933 chromosome not homologous to E. coli K-12 MG1655"
RESULT	4				complement(2198..2689)
AE005613		10592 bp	DNA		/gene="Z5334"
LOCUS	Escherichia coli O157:H7 genome, contig 3 of 3, section 232 of 290.				/complement(2198..2689)
DEFINITION	Escherichia coli O157:H7				/gene="Z5334"
ACCESSION	AE005613				/function="orf; Unknown function"
VERSION	AE005613.1				/note="No significant matches"
KEYWORDS					/codon_start=1
SOURCE	Escherichia coli O157:H7.				/transl_table=11
ORGANISM	Escherichia coli O157:H7.				/protein_id="AAG59010.1"
REFERENCE	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.				/db_xref="GI:12518688"
AUTHORS	1 (bases 1 to 10592) Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.				/translation="MKSKLQKILICLLFLCCIYNLWLRPVQILYTYSDAGNSVFLVV DHLPTWDSKINWYKLQNEIKNHPLEGSNHWYVIDINGETDYKKYIEGPEYDL YCFPTIKSNDNCITKNYLMVINEYPYRNTHTIGINDTEYQLTQENKIEQVFNPHNFK DNF"
TITLE	Genome sequence of enterohemorrhagic Escherichia coli O157:H7				complement(2686..3534)
JOURNAL	Nature 409 (6819), 529-533 (2001)				/gene="Z5335"
REFERENCE	2 (bases 1 to 10592) Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D., Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A., Posfai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,L., Grobeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamouis,K., Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C., Welch,R.A. and Blattner,F.R.				complement(2686..3534)
AUTHORS	Direct Submission Submitted (22-OCT-2000) Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA				/gene="Z5335"
TITLE					/function="orf; Unknown function"
JOURNAL					/note="No significant matches"
REFERENCE					/codon_start=1
AUTHORS					/transl_table=11

```
gene
  /protein_id="AAG59011.1"
  /db_xref="GI:12518689"
  /translation="MNTLLPSINICTLFEISLFDNYSADDMQYGDVQDDFLSLG
  LSDISAKVRLIKLTHFRYGPNGINAFSSSGTKISQRECTDILFAEMKELAKMFS
  FFOYKTLIEDLHEHRYGNFHSQQNLNLSFHEKINKYNSPIRIKCEIENGIN
  STSYQPIQLQISIKTLISSLNKFNEDFSFNGLGSVHDI SAQKISLISFOKYA
  IGWSATIHFAQDHFLDVTIDKNKTKSYKRFRIWFLQRHKDEAFKPFPTNFNTE
  RIENYL"
  3748..3867
  /gene="Z5336"
  CDS
  /gene="Z5336"
  /function="orf; Unknown function"
  /note="Residues 1 to 39 of 39 are 100.00 pct identical to
  gi16960250|gb|AAAF33440.1| (AF233324) S. typhimurium
  magnesium transport protein (CORA) (SP:P31138) [Salmonella
  typhimurium LT2]"
  /codon_start=1
  /transl_table=11
  /product="partial repeat of corA"
  /protein_id="AAG59012.1"
  /db_xref="GI:12518690"
  /translations="MNFEMPELKWSEGYPGALIFMILAGLAPLYLFKKNWL"
  complement(3892..4278)
  /gene="Z5337"
  CDS
  complement(3892..4278)
  /gene="Z5337"
  /function="orf; Unknown function"
  /note="No significant matches"
  /codon_start=1
  /transl_table=11
  /protein_id="AAG59013.1"
  /db_xref="GI:12518691"
  /translation="MFKSSIRNIFPSKILERSRFOIYIITLSWSILITLNDINLAL
  SFILVNNTVINIGAKIRIDTGCKAIYFVLISGLTLFSIIAIEGVRGESLSVQ
  LNVAILFTTMIILSAIPSAKMSLSAH"
  4483..4794
  /gene="Z5339"
  CDS
  4483..4794
  /gene="Z5339"
  /function="orf; Unknown function"
  /note="No significant matches"
  /codon_start=1
  /transl_table=11
  /protein_id="AAG59014.1"
  /db_xref="GI:12518692"
  /translation="MDSINEMVTFGSCSWKNPYSKKSDFKVKNLLWNAITLVLPIL
  SLRTTLMLRSIGLSFGFCQILGRYTMISLGSWYGVLGSPPTNIVLFLILPFKK
  R"
  complement(4841..5728)
  /gene="rard"
  CDS
  complement(4841..5728)
  /gene="rard"
  /function="orf; Unknown function"
  /note="Residues 4 to 295 of 295 are 92.12 pct identical to
  residues 8 to 299 of 300 from Escherichia coli K-12 Strain
  MG1655; B3819"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAG59015.1"
  /db_xref="GI:12518693"
  /translation="MDAKTRQCVLLAALYFIWGIAPAFKLIYVPADEILTHRVI
  WSFFVMVLMSICRWSYIKTLTQPKIFMLAVSAVLGGNMLLIWVNNHHMLEA
  SLGYFINPLVILGIFLGERFRQMLAVILAICGLVOLWTFGSLPIIALGLFAS
  FAYGLVKKRIAYEATGMLIETMULLPVAIYLPALADSPTSHMGONPMSLNLLIA
  AGIVTVPILCLCFCAATRLRLSLTGFQYIGTFLFLLAVTYGKPGADRKMTFTFI
  WVALAFVMDAILTQRRKS"
  complement(5780..6265)
  /gene="yigI"
  gene
  complement(5780..6265)
  /gene="yigI"
  /note="Z5341"
```

```
CDS
  complement(5780..6265)
  /gene="yigI"
  /function="orf; Unknown function"
  /note="Residues 1 to 161 of 161 are 100.00 pct identical
  to residues 1 to 161 of 161 from Escherichia coli K-12
  strain MG1655; B3820"
  /codon_start=1
  /transl_table=11
  /product="orf, hypothetical protein"
  /protein_id="AAG59016.1"
  /db_xref="GI:12518694"
  /translation="MKCKDDMSAVLTAEQALKLVGEMFYVHMPFNALGMELEYEKE
  FAQLAFKNQPMVMVWNAQSLILHGCVIASLDVAGLVGSLTRHETISDELQRRL
  SRMGTIDLVVDYLRPGRXERFATATSSLLRAGNKVAVARVELHNEEQLYIASATATYV
  G"
  6412..7281
  /gene="pIdA"
  /note="Z5342"
  6412..7281
  /gene="pIdA"
  /function="enzyme; Central intermediary metabolism:
  Phosphorus compounds"
  /note="Residues 1 to 289 of 289 are 100.00 pct identical
  to residues 1 to 289 of 289 from Escherichia coli K-12
  strain MG1655; B3821"
  /codon_start=1
  Query Match      95.9%; Score 592.4; DB 1; Length 10592;
  Best Local Similarity 97.4%; Pred. No. 3.9e-163;
  Matches 602; Conservative 0; Mismatches 16; Indels 0; Gaps 0;
  QY 1 atgttgatgtatttctcacgcgtgcgcattgtgcacattgtggcgttatgagccccggt 60
  Db 9307 ATGTTGATGTTATTTCACCGTCGCCANGGTGCACATTGTGGCGCTTATGAGTCCCGGT 9366
  QY 61 cccgattctttttgtctctcagaccgctgtcagtcgctccgctaaagaacgcatgatg 120
  Db 9367 CCCGATTTCCTTTTGTCTCTCAGACGCGTGTCACTCGTTCGCGTAAACAACGATGATG 9426
  QY 121 ggctgctgggcattacctcgcgcgtaagtgttggcgtgggatgcgctgcttggcctg 180
  Db 9427 GCGCTGCTGGGTATTACCTCGCGCGTTATGGTTTGGCGAGGATTGCGTTGGCGCTG 9486
  QY 181 catttgattatcgaaaaatggcctggctgcatacgtcattgattgctggcggtgctgctg 240
  Db 9487 CATTTGATTATCGAAAAAATGGCTGGCTGCATACGCTGATTTATGTTGCGGTGGCCTG 9546
  QY 241 tatctctgctggatgggtaccagatgctacgtggtgcactgaaaaaagagcggtttct 300
  Db 9547 TATCTCTGCTGGATGGGTTTACCAGATGCTCGCGGCGCGCTGAAAAAAGAGCGGTTCT 9606
  QY 301 gcacctgcgcacaggtcgagctggcgaaaagtggcgcgagtttctctgaaagtttactg 360
  Db 9607 GCACCTGGCGCCACAGTCTGAGCTGGCGGCGGCGAGCTTCTCTGAAAGGTTTACTG 9666
  QY 361 accaatctgcgtaatccgaaagcattatctactctggcgtgggtgtctcattgtttgtc 420
  Db 9667 ACCAATCTCGTAAATCCGAAAGCGATTATCTACTTTGGCTCGGTGTTCTCATTTGTTGC 9726
  QY 421 ggtataaactgtgcactacgcgcgctgggcatttttgcgctatcatcattcgaacg 480
  Db 9727 GGTGATACGTTGGCACTACCGCGCGCTGGGCGATTTTTCGGTTGATTATTGTCGAACG 9786
  QY 481 ctggcgtggtttaccctcgctgtgcacgctgtttgcgcctgcccaaatgcgcgctggttat 540
  Db 9787 CTGGCGTGGTTTACTGTGTGGCGCAGCGCTGTTTGGCCCTGCCGCAATGCCGCGTGTAT 9846
  QY 541 caacgtctgcgaagtgaattgatttttgcggcgaccttatttccgattttgcgatt 600
  Db 9847 CAACGCTCGGCGAAGTGATGATGATGTTTTCGCGGCGGCTTATTTCCCGGATTTGCAATT 9906
  QY 601 catttgattatttcgcgg 618
  |||||
```

```
Db 9907 CATTGATTATTTCGGG 9924
RESULT 5
AC073768/c
LOCUS AC073768 279589 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone Rp23-359N18, WORKING DRAFT SEQUENCE, 60
unordered pieces.
ACCESSION AC073768
VERSION AC073768.1 GI:8810385
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 279589)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 279589)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 1875058
Center clone name: RPCI-23_359N18
-----
Summary Statistics
Consensus quality: 230709 bases at least Q40
Consensus quality: 233950 bases at least Q30
Consensus quality: 258916 bases at least Q20
Estimated insert size: 216000; agarose-fp estimation
Insert coverage: 7.54 in Q20 bases; agarose-fp estimation
Quality coverage: 9.53 in Q20 bases; sum-of-contigs estimation.
NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
*
* 1009: contig of 1009 bp in length
* 1010: gap of unknown length
* 1109: contig of 1023 bp in length
* 2132: gap of unknown length
* 2133: contig of 1025 bp in length
* 2232: gap of unknown length
* 3257: contig of 1193 bp in length
* 3358: 4550: contig of 1193 bp in length
* 3358: 4550: gap of unknown length
* 4651: 4614: contig of 1764 bp in length
* 6415: 6514: gap of unknown length
* 6515: 7557: contig of 1043 bp in length
* 7658: 8885: contig of 1228 bp in length
* 8886: 8985: gap of unknown length
* 8986: 10366: contig of 1381 bp in length
* 10367: 10466: gap of unknown length
* 11691: 11691: contig of 1225 bp in length
* 11692: 11791: gap of unknown length
* 11792: 13116: contig of 1325 bp in length
* 13117: 13216: gap of unknown length
* 13217: 14234: contig of 1018 bp in length
* 14235: 14334: gap of unknown length
* 14335: 15472: contig of 1138 bp in length
* 15473: 15572: gap of unknown length
*
15573
16690
16789: contig of 1117 bp in length
16789: gap of unknown length
17838: contig of 1049 bp in length
17838: gap of unknown length
17939: contig of 1250 bp in length
19188: gap of unknown length
19288: gap of unknown length
20826: contig of 1538 bp in length
20926: gap of unknown length
21984: contig of 1058 bp in length
22084: gap of unknown length
23158: contig of 1074 bp in length
23258: gap of unknown length
23259: contig of 1339 bp in length
24697: gap of unknown length
25784: contig of 1087 bp in length
25884: gap of unknown length
26929: contig of 1045 bp in length
27029: gap of unknown length
28535: contig of 1506 bp in length
28635: gap of unknown length
30028: contig of 1393 bp in length
30228: gap of unknown length
31331: contig of 1203 bp in length
31431: gap of unknown length
32511: contig of 1080 bp in length
32611: gap of unknown length
33889: contig of 1278 bp in length
33989: gap of unknown length
35192: contig of 1203 bp in length
35292: gap of unknown length
36574: contig of 1282 bp in length
36774: gap of unknown length
37945: 38044: contig of 1270 bp in length
37945: 38044: gap of unknown length
39129: contig of 1085 bp in length
39229: gap of unknown length
40571: contig of 1342 bp in length
40671: gap of unknown length
42172: contig of 1501 bp in length
42272: gap of unknown length
44167: contig of 1895 bp in length
44267: gap of unknown length
45266: contig of 1459 bp in length
45826: gap of unknown length
47489: contig of 1663 bp in length
47589: gap of unknown length
48878: contig of 1289 bp in length
48978: gap of unknown length
50766: contig of 1788 bp in length
50866: gap of unknown length
54337: contig of 3371 bp in length
54337: gap of unknown length
56520: contig of 2183 bp in length
56620: gap of unknown length
58133: contig of 1513 bp in length
58233: gap of unknown length
59527: contig of 1294 bp in length
59627: gap of unknown length
63467: contig of 3840 bp in length
63567: gap of unknown length
65727: contig of 2160 bp in length
65827: gap of unknown length
68787: contig of 2960 bp in length
68987: gap of unknown length
72316: contig of 3429 bp in length
72416: gap of unknown length
76575: contig of 4159 bp in length
76675: gap of unknown length
82272: contig of 5597 bp in length
82372: gap of unknown length
86199: contig of 3827 bp in length
86299: gap of unknown length
91846: contig of 5547 bp in length
86300
```



```
* 91847 91946: gap of unknown length
* 91947 96394: contig of 4448 bp in length
* 96395 96494: gap of unknown length
* 96495 102483: contig of 5989 bp in length
* 102484 102583: gap of unknown length
* 102584 110476: contig of 7893 bp in length
* 110477 110576: gap of unknown length
* 110577 119241: contig of 8665 bp in length
* 119242 119341: gap of unknown length
* 119342 130059: contig of 10718 bp in length
* 130060 130159: gap of unknown length
* 130160 142204: contig of 12045 bp in length
* 142205 142304: gap of unknown length
* 142305 157417: contig of 15113 bp in length
* 157418 157517: gap of unknown length
* 157518 185392: contig of 27875 bp in length
* 185393 185492: gap of unknown length
* 185493 212309: contig of 26817 bp in length
* 212310 212409: gap of unknown length
* 212410 249959: contig of 37550 bp in length
* 249960 250059: gap of unknown length
* 250060 279589: contig of 29530 bp in length.
FEATURES
  Location/Qualifiers
    1..279589
      /organism="Mus musculus"
      /db_xref="taxon:10090"
      /clone="RP23-359N18"
      /clone_lib="RPCI mouse BAC library 23"
BASE COUNT 69254 a 68603 c 67231 g 68544 t 5957 others
ORIGIN

Query Match      90.2%; Score 557.2; DB 75; Length 279589;
Best Local Similarity 98.7%; Pred. No. 1..18-152;
Matches 614; Conservative 0; Mismatches 3; Indels 5; Gaps 5;

QY 1 agttgatgtttttctcaaccgtcccatggtgcacattgtggccttatagcccggt 60
Db 30830 ATGTGATGTTATTTCTACCGTCGCCATGTGTCACATTGTGGCGTATGAGCCCGGT 30771

QY 61 cccgattctttttgtctctcagaccgctgtcagctgcattcccgtaagaagcagatg 120
Db 30770 CCGATTTCTTTTGTCTCTCAGACCCTGTCAGTCGTTCCTCCGTAAGAAGCCGATG 30711

QY 121 ggcgtgctggcatcaccctgcggcgaatggtttggctgggattgcctcttggcctg 180
Db 30710 GCGTGTCTGGCATACCTCGCGCGTATGTTGGCTGGGATGCGCTGCTGGCCTG 30651

QY 181 catttgattatcgaaaaaatggcctggctgcatacgcgtgattatggtggcggtgctg 240
Db 30650 CATTTTGATTATCGAAAAATGGCCTGGCTGCATACGCTGATTATGTTGGCGGTGGCCTG 30591

QY 241 tatctctctgtaggtttaccagatgctacgtgtgctcactgaaagagag-cggtttc 299
Db 30590 TATCTCTGCTGATGGTTTACCAGATGCTACGTGTGGTGCACGTAAGAAAGAGCCGGTTTC 30531

QY 300 tgcacctgcgcacaggtgcagctggcgaaaaagtggcgagtttccctgaaagtttact 359
Db 30530 TGCACCTGCGCCACAGTGCAGCTGGCGAAAAAGTGGCGCAGTTTCCCTGAAAGGTTTACT 30471

QY 360 gaccatctcgctaatccgaaagcattatctacttgcgtcgattctctcattgtttgt 419
Db 30470 GACCATCTGCTGATATCCGAAAGCAGATATCTPACTTTCGCTCGGTGTCTCTCATTTGTTGT 30411

QY 420 cgggtgataacgttggcactaccgcgcgtggggcaatttttgcgtgatca-ttgcgaaa 478
Db 30410 CGGTGATAACGTTGGCACATACCAGCGCGTGGGGCAATTTTGGCGTGATCATCTTGTGCGAAA 30351

QY 479 cgtgcgcgtggtttaccgtcgtt-gccagcctgt-ttgcctgcgcgaaatgcgccgtg 536
Db 30350 CGCTGCGGTGTTTACCGTCGTTGCCAGCCGTCTTGCCTTGCAGCAATGCGCCGTGG 30291

QY 537 ttatcaacgctctggcgaaagtggattgattgttttccggggcggttatttgcgcgattgg 596
```

```
Db 30290 TTATCTGCGGAAGTGGATTGATGTTTTCGCGGCGTATTTCGCGGATTGG 30232
QY 597 cattcatttgatttttcgcgg 618
Db 30231 CATTCATTGATTATTTCGCGG 30210

RESULT 6
AC020970 212936 bp DNA HTG 10-FEB-2000
LOCUS Mus musculus clone RP23-252M21, WORKING DRAFT SEQUENCE, 144
DEFINITION unordered pieces.
ACCESSION AC020970
VERSION AC020970.1 GI:6691260
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 212936)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 212936)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Estimated insert size: 212936; sum-of-contigs estimation
Estimated insert size: 157000; pulse field gel estimation
Quality coverage: 4.20x in Q20 bases; pulse field gel estimation
Quality coverage: 3.10x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 144 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1 1159: contig of 1159 bp in length
* gap of unknown length
* 1160 2363: contig of 1204 bp in length
* gap of unknown length
* 2364 3604: contig of 1241 bp in length
* gap of unknown length
* 3605 5188: contig of 1584 bp in length
* gap of unknown length
* 5189 6372: contig of 1184 bp in length
* gap of unknown length
* 6373 7737: contig of 1365 bp in length
* gap of unknown length
* 7738 9783: contig of 2046 bp in length
* gap of unknown length
* 9784 11046: contig of 1263 bp in length
* gap of unknown length
* 11047 12223: contig of 1177 bp in length
* gap of unknown length
* 12224 13492: contig of 1269 bp in length
* gap of unknown length
* 13493 14669: contig of 1177 bp in length
* gap of unknown length
* 14670 15886: contig of 1217 bp in length
* gap of unknown length
* 15887 17089: contig of 1203 bp in length
* gap of unknown length
```



```
OY 61 cccgatttcttttttctcagaccgctcagtcggttcctcgttaagcagcgatgatg 120
Db 174321 CCCGATTC-TTTTTGTCCTCACACCGCTCGTCASTCGTCCCGTAAGAGCGATGAT- 174378

OY 121 ggcgtcgggcattacacgcgcgtgaatggtttggcgtgggattgcgctgtgcctg 180
Db 174379 GCGTGCCTGGGCTATACCTGGCGGCTAATGTTTGGCTGGGATTCGCT-CTTGGCCTG 174437

OY 181 catttgattatcgaaataagcctgctgcatacgcgtgattatggtggcggtgcctg 240
Db 174438 CATTTGATTATCGAAAAATGCCCTGGCTGCATACGCTGA-TATGGTGGCGGTGGCCTG 174496

OY 241 tatctctcgtggtggtttaccagatgctacgtggtgcactgaaagagcggtttct 300
Db 174497 TATCTCTGCTGATGGGTTACAGATGCTA-GTGGTGCACCTGAAAAAGAGCGGTTCT 174555

OY 301 gcaactggccacagtgctgaactggcgaaagtgggcgagtttctcgtgaaggttaactg 360
Db 174556 GCACCTGGCCACAGGTGGA-CTGGCGAAAAGTGGCGCAGTTTCTTGAAAGGTTTACTG 174614

OY 361 accaatctcgcttaaccgaaagcgtattctactttgctcgggtgttctcattgtttgtc 420
Db 174615 ACCAATCTCG-TAATCCGAAAGCATTATCTACTTTGCTCGGTGTTCTCATTTGTTGTC 174673

OY 421 ggtgataacgttggcactacgcgcgtggtgggcaatttttgcgctgcatctgcgaacg 480
Db 174674 -GTGATAACGTTGGCATTACCGCGCGCTGGGGCATTTTGGCTGTGATCA-TGTCGAACG 174731

OY 481 ctggcgtggtttaccgtgttgccagcgtgttgcctgcgcgaatgcgcctggttat 540
Db 174732 CTGGCGTGGTTACGCTGTTGCCAGCTGTTTGCCCTG-CGCAAAATGCGCGGTGTTAT 174790

OY 541 caactctggcgaagtgatgattggttttgcggggcggttattgcgcatgttgcatt 600
Db 174791 CAACGCTCGGCAAGTGGATGGTGGTTT-CCGGGGCGTATTATTCGCGGATTGGCAAT 174849

OY 601 catttgattatttcqcg 618
Db 174850 CATTTGATTATTTCGCG 174867

RESULT 7
STYSTD1
LOCUS STYSTD1 96086 bp DNA BCT 11-FEB-2000
DEFINITION Salmonella typhimurium fragment STMD1.
ACCESSION AF233324
VERSION AF233324.1 GI:6960215
KEYWORDS
SOURCE Salmonella typhimurium LT2.
ORGANISM Salmonella typhimurium LT2
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
1 (bases 1 to 96086)
AUTHORS Washington University Genome Sequencing Center.
TITLE The Salmonella typhimurium Genome Sequencing Project
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 96086)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: sclifton@watson.wustl.edu or
jspieth@watson.wustl.edu
```

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>= 30); an attempt was made to resolve all sequencing problems,

such as compressions and repeats; all regions were covered by
sequence from more than one ml3 subclone.

NOTES:

Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs
GLIMMER (Salzberg, S. Delcher, A., Kasif, A. and White, O. (1998)
NAR 26,544-548), and GeneMark (Lukashin, A.V. and Borodovsky, M.
(1998), NAR 26,1107-1115.

```
Location/Qualifiers
1. .96086
/organism="Salmonella typhimurium LT2"
/strain="SGSC1412"
/db_xref="taxon:99287"
/clone="STMD1"
<1. .980
/note="rrlC"
/product="23S ribosomal RNA"
1072. .1191
/note="rrfC"
/product="5S ribosomal RNA"
1244. .1323
/product="trNA-Asp"
1329. .1401
/product="trNA-Trp"
complement(1506. .2354)
/gene="STMD1.99"
complement(1506. .2354)
/gene="STMD1.99"
/note="83% identity with E. coli hypothetical protein  
(YifA) (SP:P22788) and 95% identity with amino acids 1-80  
of E. coli possible regulatory protein (PSSR) (SP:P27826)."
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAF33488.1"
/db_xref="GI:6960298"
/translation="MESTVDTLLKTFLEVSPTHFGRAAEALYLTVSAVSFRIROLE  
NQLGMLFRRNRNRLTTAGKLLPYAETLNTWQAAKEVAHTSRINERSIGASAS  
LWECMLNAGRLYQLQEPQSQFQFARQAQSLVKOLHQRQLDLITTPAKMDEF  
SSQLGHFTALYCSPARKKSELNLYRLWGPDPQOQHEGLIAADEVPVLTTSSEL  
ARQLSALINGCSWLPVNWANEKGLHTVADSATLSRPLYAIWLQNSDKYSLICDLKT  
DVLDEQ"
2461. .2799
/gene="yifE"
2461. .2799
/gene="yifE"
/note="STMD1.96"
/codon_start=1
/transl_table=11
/product="95% identity with E. coli hypothetical protein  
(YifE) (SP:P27827)"
/protein_id="AAF33485.1"
/db_xref="GI:6960295"
/translation="MAESFTTNRYFDNKHYPGRFSRHGDTFKEAQLLEHRHAFND  
LDLGKRPVTEEEKLFVAVCGEREPTDAERVWSKYMTRIKPKRFTLLSGKRPQE  
GAEDYTEADD"
complement(2824. .4344)
/gene="yifB"
complement(2824. .4344)
/gene="yifB"
/note="STMD1.95"
/codon_start=1
/transl_table=11
/product="85% identity to E. coli hypothetical protein  
(YifB) (SP:P22787); contains similarity to Pfam family  
PF01078 (Magnesium chelatase, subunit ChlI), score=351.8,  
E=7.6e-102, N=1"
/protein_id="AAF33484.1"
/db_xref="GI:6960294"
/translation="MSLAIVHTRAAALGVNAPPTITIEVHISNGLPGLTMVGLPETTYKE
```

FEATURES
source

rRNA

rRNA

tRNA

tRNA

gene

CDS

gene

CDS

gene

CDS

```

gene
  7948..9798
  /gene="ilvD"
CDS
  7948..9798
  /gene="ilvD"
  /note="STMd1.88"
  /codon_start=1
  /transl_table=11
  /product="S typhimurium dihydroxy-acid dehydratase (ilvD)
  (SP:P40810); contains similarity to pfam family PF00920
  (Dehydratase family). score=1023.9, E=3.5e-304, N=1"
  /protein_id="AAF33480.1"
  /db_xref="GI:6960290"
  /translation="MPKYRSATVTHGRNNAGARALWRATGMTDSDFGPIIAVNSFT
  QFVGHVHLRDLGLKVAEQIEASGVAKFEFNTIAVDDGIAMGHGSMYSLPSRELIAD
  SVEYVNAICADAMVCISNCDKITPGLMASLRINPFIYFVGGPMEAGTKLSDKII
  KLDLVDMATQADPKVSDDSQNSQVRSACPTGCGSCGMFTANSMNCLFEALGLSQPGN
  GSLLATHADRKOLFNAKRIIVELTKRYVEQNDESALPRNIAASKAAFNAMTLDIAMG
  GSTNTVLHLAAQEAEDFTMSDIDKLSKVPOLCKVAPSTOKYHMEDEVHAGGVLG
  ILGELDRLGLNRNVKVLGLTLPQTLQEQYDITVTQDEAVKMFRAAGAGRTQAES
  QCRWDSLDDRAAGCIRSEYASKDGGLAVLYGNFAENGCIKRTAGSDVSLKFTG
  PAKVYESQDDAVEAILGKRVGVDVVIYRIGPKGPGQEMLYPTSEKSMGLKAC
  ALITDGRFSGTSGLSIGHVSPAAAGGTIALIEDGDTIAIDIPNRSLQLSPAETI
  ARREAQARGDKAWTPKNRQGVSPALRAYASLATSDKGAVRDKSLKG"
  9801..11345
  /gene

Query Match      71.5%; Score 442.6; DB 3; Length 96086;
Best Local Similarity 82.3%; Pred. No. 4.7e-119;
Matches 508; Conservative 0; Mismatches 109; Indels 0; Gaps 0;

Qy 1 atgtgatgtattctcacccgtcccatggtgcacattgtggcgtttatgagcccgat 60
Db 60428 ATGATGATGCTATTTTTCACCGTGGCAATGGTGACATGTTGCGCTAATGACCCCGC 60487

Qy 61 ccgattctcttttctctcagaccgctgtcagtcgtcccgtaaaagacgatgatg 120
Db 60488 CCGACTCTTTCTCTGCTGCTCAAACTGCTGTCAGCGCTTCGCTAAAGAACGATGATG 60547

Qy 121 ggcgtcgtggcattacctgcgcgctaagtgttggcgtggatgctgcctgctgtgacctg 180
Db 60548 GCGCTACTGGGCATCACTCGCGCGTGATGGTATGGCGGCGCTGCGCTGCTGCGCTG 60607

Qy 181 catttgattatcaaaaaatggcctggctgcatacctgattatggtggcggtggcctg 240
Db 60608 CATCTCATCATGAAAAATGGCTGGCTGCATACGATCATTTATGTTAGCGCGCGCTG 60667

Qy 241 tatctctcgtggatgggtttaccagatgctacgtggtgcactgaaaaagagcggtttct 300
Db 60668 TACCTGTCTGGATGGCTATCAGATGCTGCGCGCGCGCTGAAAAACAGGATCGCGG 60727

Qy 301 gcacctgcgcacagctcagctggtgcgaaaagtggcgagttcttcgaaagtgttactg 360
Db 60728 GCGTCTTCTCCGCACATTTGAACCTGGCGCAGAGCGCGCGAGCTTCTCAAAAGGCTGTG 60787

Qy 361 accaactcgtcgaatccgaaagcattatctcttggctcgtggttctctcattgtttgc 420
Db 60788 ACCAATCTGCGAATCCTTAAAGCGATATCTATTATTTGGTTCGCTTTTTCACCTTTTGC 60847

Qy 421 ggtgataacggttggcactaccgcgcgtggggaatttttggcgtgatcatattgcgaaag 480
Db 60848 GCGGATAAAGTCGCGCGCTGCGCGCGCTGGGGTATTTTTCGCGTAAATACCTCGGAACG 60907

Qy 481 ctggcgtggtttaccgctgttgcagcctgttgcctgcgcgaatgcgcgctggttat 540
Db 60908 CTGGCTGTTTACCGTATGTCGCCACCTGTTCGCGCTGCCGAAATATGCGCGCTGGCTAT 60967

Qy 541 caacgtctggcgaagtggattgatggttttgcggggggttatttggcgggatttggcatt 600
Db 60968 CAGCGCTGCGGAAATGGATGATGCTTTTGGCGGCTTTCGCGGCTCTCTTTCGCGGCTTTGTGATT 61027

Qy 601 catttgattatttcgcg 617
Db 61028 CACCTGATTATTTCGCG 61044
```

```

ADRVRSALINGSVEFFPAKTTIINAPADLPKEGGRYDLPIAVALLAASBOLTASNLE
AYELVGLATGALRGVPGATSSATEAIRAGRNIIIVATENAAEVLISKEGCFIADHL
QTVCAFLGHALERPLAQMASATADLREDVIGORQKRGLEITRAAGHNLILIGP
PTGCTKMLASGILPLSNEEALESAILSLVNADTVOKRQORFPRSPHBSASLT
AWGCGAIPAPGEISLAHNGILFJDELPFEFERTLDALREPIESGQIHLSTRAKITY
PAROLIAAMNSPTGYQGNHNRCTPEQTLRYLNRLSGPFLDFDLSEIPLPPPGI
LSQHASGESASPTVKRYIAAHERQYRQKKLNARLEGREIQKYCVLHDDARWLEDT
LVHGLSTRAWORLLKARTIADIELAQISOIHSROHLQEAUSYRAIDRULLIHLQKLLA"
4696..4794
  /gene="ilvL"
  /gene="ilvL"
  /note="STMd1.92"
  /codon_start=1
  /transl_table=11
  /product="100% identity to E. coli ilvGMDA operon leader
  peptide (ilvL) (SP:P03060)"
  /protein_id="AAF33486.1"
  /db_xref="GI:6960296"
  /translation="MTALLRVISLVVISVVVLIIPPCGAALGRKA"
  4934..6580
  /gene="ilvG_1"
  /gene="ilvG_1"
  /note="STMd1.91"
  /codon_start=1
  /transl_table=11
  /product="92% identity with E. coli acetolactate synthase
  II (ilvG) (SP:P08992); contains similarity to pfam
  family PF00205 (Thiamine pyrophosphate enzymes),
  score=952.6, E=1.5e-295, N=1"
  /protein_id="AAF33483.1"
  /db_xref="GI:6960293"
  /translation="MNSAQVNVHRAQGVKTVFGYPGGAIMPVYDALYDGVVEHLIC
  RHEGAAWAIIGTARSTKGTGVCVATNITGLADALDSVPPVAVITGVQVAP
  FIGDAFQEVDFVLGSLACTKHSPLQSLLEELPRIMAEAFEVANRPGVPLVIDPKD
  IQLASGEVWFTTVANEATFPQADVEAQRLMEQAKPMLYGGVGYMAQVPAIRLK
  FIAYTQMPVCTTKLGADEADYPYLCMLGMHGTKAANFAVQECOLLIAVAGARFDR
  VTGLNTFAPNASVTHMDIDPAEMNKLRQAHVALQGLNSLLPALQOPLKIDAWROSC
  AELRAEHAWYDHGETIYAPLLKOLSERKPADSVVTTDVGOHOMWSAOMHTYTRPE
  IFTSSGLGPMGFCLPAANGAQAARPNDTVICISGDGSMNVYQELGTVYRKQPLPKI
  VLLDNQRLGWVRQWQQLFFQERYSETLTNDPFLMASAFGIPGQHITRKQDOVEAL
  DTLMASGYPYLHVSIDELENVPLVPFGASNSEMLEKLS"
  6580..6840
  /gene="ilvM"
  /gene="ilvM"
  /note="STMd1.90"
  /codon_start=1
  /transl_table=11
  /product="93% identity with E. coli acetolactate synthase
  I (ilvM) (SP:P13048)"
  /protein_id="AAF33482.1"
  /db_xref="GI:6960292"
  /translation="MOHOVNVSAFNPETLRLVVRVHRFGVCSNMNEAADAQNI
  NIELTVASPRSDLLFSQLSKLDVVAHVACQSAATSQQIRA"
  6858..7787
  /gene="ilvE"
  /gene="ilvE"
  /note="STMd1.89"
  /codon_start=1
  /transl_table=11
  /product="S. typhimurium branched-chain-amino-acid
  transaminase (ilvE) (SP:P15168)"
  /protein_id="AAF33481.1"
  /db_xref="GI:6960291"
  /translation="MTTKADYIWFNGEMVRWEDAKVHVMSHALHYGTGSVFEGIRCYD
  SHKGPVVRHREHMRDLSAKIYRFPVSQSIDELMAKCRDVIKKNLTSAYIRPLVF
  VGDVGMGNPPPGYTTDVIITAAPWGAYLGAELDQCIDAMVSNWRAPNPTPAK
  AGNYLSLLVAGEARHGVEGLADVNGYISEGAGENLFYKDGVLFPPTTSAL
  PGITRDALIKIARELGEVREQVLSRESLYLADEVFMSGTAAEITPVRSVDGIQVGE
  RCGPVTARIQQAFFGLFTGETDKWGLDPVNS"
```

RESULT 8
AC020885/c
LOCUS AC020885 268294 bp DNA HTG 16-FEB-2000
DEFINITION Mus musculus clone RP23-46411, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC020885
VERSION AC020885.2 GI:6980212
KEYWORDS HTG; HTGS.PHASE0.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 268294)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 268294)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Feb 16, 2000 this sequence version replaced gi:6686423.
* NOTE: This record contains 183 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
* 1 1013: contig of 1013 bp in length
* gap of unknown length
* 1014 1615: contig of 602 bp in length
* gap of unknown length
* 1616 2405: contig of 790 bp in length
* gap of unknown length
* 2406 3234: contig of 829 bp in length
* gap of unknown length
* 3235 3533: contig of 299 bp in length
* gap of unknown length
* 3534 4467: contig of 934 bp in length
* gap of unknown length
* 4468 5401: contig of 934 bp in length
* gap of unknown length
* 5402 6154: contig of 753 bp in length
* gap of unknown length
* 6155 6629: contig of 475 bp in length
* gap of unknown length
* 6630 7176: contig of 547 bp in length
* gap of unknown length
* 7177 7824: contig of 648 bp in length
* gap of unknown length
* 7825 8656: contig of 832 bp in length
* gap of unknown length
* 8657 8865: contig of 209 bp in length
* gap of unknown length
* 8866 9544: contig of 679 bp in length
* gap of unknown length
* 9545 10280: contig of 736 bp in length
* gap of unknown length
* 10281 10975: contig of 695 bp in length
* gap of unknown length
* 10976 11742: contig of 767 bp in length
* gap of unknown length
* 11743 12178: contig of 436 bp in length
* gap of unknown length
* 12179 12887: contig of 709 bp in length
* gap of unknown length
* 12888 13607: contig of 720 bp in length
* gap of unknown length
* 13608 13911: contig of 304 bp in length
* gap of unknown length
* 13912 14819: contig of 908 bp in length
* gap of unknown length
* 14820 15571: contig of 752 bp in length
* gap of unknown length
* 15572 15872: contig of 301 bp in length
* gap of unknown length
* 15873 16608: contig of 736 bp in length
* gap of unknown length
* 16609 16851: contig of 243 bp in length
* gap of unknown length
* 16852 17277: contig of 426 bp in length
* gap of unknown length
* 17278 18057: contig of 780 bp in length
* gap of unknown length
* 18058 18704: contig of 647 bp in length
* gap of unknown length
* 18705 19414: contig of 710 bp in length
* gap of unknown length
* 19415 20372: contig of 958 bp in length
* gap of unknown length
* 20373 21010: contig of 638 bp in length
* gap of unknown length
* 21011 21258: contig of 248 bp in length
* gap of unknown length
* 21259 22363: contig of 1105 bp in length
* gap of unknown length
* 22364 23639: contig of 1276 bp in length
* gap of unknown length
* 23640 24853: contig of 1214 bp in length
* gap of unknown length
* 24854 25795: contig of 942 bp in length
* gap of unknown length
* 25796 27027: contig of 1232 bp in length
* gap of unknown length
* 27028 27729: contig of 702 bp in length
* gap of unknown length
* 27730 28722: contig of 993 bp in length
* gap of unknown length
* 28723 29714: contig of 992 bp in length
* gap of unknown length
* 29715 30858: contig of 1144 bp in length
* gap of unknown length
* 30859 31397: contig of 539 bp in length
* gap of unknown length
* 31398 31671: contig of 274 bp in length
* gap of unknown length
* 31672 32959: contig of 1288 bp in length
* gap of unknown length
* 32960 33791: contig of 832 bp in length
* gap of unknown length
* 33792 34848: contig of 1057 bp in length
* gap of unknown length
* 34849 35060: contig of 212 bp in length
* gap of unknown length
* 35061 35776: contig of 716 bp in length
* gap of unknown length
* 35777 35903: contig of 127 bp in length
* gap of unknown length
* 35904 36602: contig of 699 bp in length
* gap of unknown length
* 36603 37110: contig of 508 bp in length
* gap of unknown length
* 37111 38059: contig of 949 bp in length
* gap of unknown length
* 38060 38730: contig of 671 bp in length
* gap of unknown length
* 38731 39790: contig of 1060 bp in length
* gap of unknown length
* 39791 40327: contig of 537 bp in length
* gap of unknown length
* 40328 40442: contig of 115 bp in length

*	40443	41688:	contig of 1246 bp in length	gap of unknown length
*			gap of unknown length	
*	41689	42688:	contig of 1000 bp in length	gap of unknown length
*			gap of unknown length	
*	42689	43847:	contig of 1159 bp in length	gap of unknown length
*			gap of unknown length	
*	43848	44511:	contig of 664 bp in length	gap of unknown length
*			gap of unknown length	
*	44512	45780:	contig of 1269 bp in length	gap of unknown length
*			gap of unknown length	
*	45781	47202:	contig of 1422 bp in length	gap of unknown length
*			gap of unknown length	
*	47203	48647:	contig of 1445 bp in length	gap of unknown length
*			gap of unknown length	
*	48648	49652:	contig of 1005 bp in length	gap of unknown length
*			gap of unknown length	
*	49653	50485:	contig of 833 bp in length	gap of unknown length
*			gap of unknown length	
*	50486	51527:	contig of 1042 bp in length	gap of unknown length
*			gap of unknown length	
*	51528	53119:	contig of 1592 bp in length	gap of unknown length
*			gap of unknown length	
*	53120	53623:	contig of 504 bp in length	gap of unknown length
*			gap of unknown length	
*	53624	54669:	contig of 1046 bp in length	gap of unknown length
*			gap of unknown length	
*	54670	55309:	contig of 640 bp in length	gap of unknown length
*			gap of unknown length	
*	55310	56382:	contig of 1073 bp in length	gap of unknown length
*			gap of unknown length	
*	56383	56625:	contig of 243 bp in length	gap of unknown length
*			gap of unknown length	
*	56626	57345:	contig of 720 bp in length	gap of unknown length
*			gap of unknown length	
*	57346	57505:	contig of 160 bp in length	gap of unknown length
*			gap of unknown length	
*	57506	58908:	contig of 1403 bp in length	gap of unknown length
*			gap of unknown length	
*	58909	59968:	contig of 1060 bp in length	gap of unknown length
*			gap of unknown length	
*	59969	61377:	contig of 1409 bp in length	gap of unknown length
*			gap of unknown length	
*	61378	62291:	contig of 914 bp in length	gap of unknown length
*			gap of unknown length	
*	62292	63319:	contig of 1028 bp in length	gap of unknown length
*			gap of unknown length	
*	63320	63818:	contig of 499 bp in length	gap of unknown length
*			gap of unknown length	
*	63819	64823:	contig of 1005 bp in length	gap of unknown length
*			gap of unknown length	
*	64824	65269:	contig of 446 bp in length	gap of unknown length
*			gap of unknown length	
*	65270	65904:	contig of 635 bp in length	gap of unknown length
*			gap of unknown length	
*	65905	66815:	contig of 911 bp in length	gap of unknown length
*			gap of unknown length	
Query Match 67.3%; Score 415.8; DB 66; Length 268294;				
Best Local Similarity 99.5%; Pred. No. 3.8e-111;				
Matches 417; Conservative 0; Mismatches 2; Indels 0; Gaps 0;				
QY	200	tggcctggctgcatacgctgattatgggcggtggcggtgctgtatctgctgtaggggtt	259	
Db	11742	TGCCCTGGCTGCATACGCTGATATGGTGGCGGTGGCCCTGTATCTCTCGATGGGTT	11683	
QY	260	accagatgctacggtgcactgaaaaagagcggtttctgcacctgcgcacaggtcgc	319	
Db	11682	ACCAATGCTACGTGGTGGCTAGAAAGAGGCGGTTTCTGCACCTGCGCCACAGTTCG	11623	
QY	320	agctggcgaaagtggcgagttctctgaaagggtttactgaccaatctcgctaataccga	379	
Db	11622	AGCTGGCGAAAAGTGGCGCAGTTTCTCTGAAAAGGTTTACTGACCAATCTCGCTAATCCGA	11563	

QY	380	aagcgattatctacttggctgctggtgttcttcattgttttgcgtgataaacgttggcacta	439	
Db	11562	AAGCGATTATCTACTTTGGCTGGTGTCTCAATGTTTTTGTGCGGTGATAACGTTGGCACA	11503	
QY	440	cgcgcgctggggcatttttgcgctgctgctgctgctgctgctgctgctgctgctgctg	499	
Db	11502	CCGCGCGCTGGGGCATTGTCGCTGATCATTTGCGAAAGCGTGGCTGTTTACCGTCG	11443	
QY	500	ttgccagctgtttgcccctgcgcgcaaatgcgcgtggtttatcaacgtctgctgcaagtgg	559	
Db	11442	TTGCCAGCCTGTTTGGCCTGCGCAATGCGCGTGTGTTATCAACGCTGTCGGCGAAGTGA	11383	
QY	560	ttgatggtttgcggggcgatttttcggtgatttgcgctgctgctgctgctgctgctgctg	618	
Db	11382	TTGATGTTTGGCGGGCGCTTATTTGCCGGATTGGCATTTCATTGATTATTTTCGCGG	11324	

RESULT 9

AC079167 290452 bp DNA HTG 30-AUG-2000

LOCUS Mus musculus chromosome 6 clone RP23-226D23 strain C57BL6/J, ***

DEFINITION SEQUENCING IN PROGRESS ***, 196 unordered pieces.

AC079167

AC079167.2 GI:9945047

KEYWORDS HTG; HTGS PHASE1.

SOURCE house mouse.

ORGANISM Mus musculus

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE 1 (bases 1 to 290452)

AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R., Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M., Goltz,J.S. and Kucherlapati,R.

TITLE High Throughput Mouse Sequencing

JOURNAL Unpublished

REFERENCE 2 (bases 1 to 290452)

AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R., Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A., Gordon,M., Goltz,J.S. and Kucherlapati,R.

TITLE Direct Submission

JOURNAL Submitted (23-AUG-2000) Department of Molecular Genetics, Albert Einstein College of Medicine Genome Center, 1300 Morris Park Ave., Bronx, NY 10461, USA

COMMENT On Aug 30, 2000 this sequence version replaced gi:9885883.

-----Genome Center Center: Albert Einstein College of Medicine Center

Code: AECOM

Web site: http://sequence.aecom.yu.edu/cgi-bin/ws.exe/mouseDB/mouseSeq/mouseSeqtable.hts

Contact: jhan@sequence.aecom.yu.edu

-----Summary Statistics

Center project name: AEN

Sequencing vector: pUC18; L08752

Chemistry: Dye-terminator Big Dye; 100%

*Consensus quality: 196511 at least Q20

*Consensus quality: 172308 at least Q30

*Consensus quality: 138790 at least Q40

Estimated insert size: agarose-FP - N/A

**Estimated insert size: 286552 - sum-of-contigs

Quality coverage: agarose-FP - N/A

Quality coverage: 3.3x sum-of-contigs - N/A

* NOTE: This is a 'working draft' sequence. It currently consists of 196 contigs. The true order of the pieces is not known and their order in this sequence record is arbitrary. Gaps between the contigs are represented as runs of N, but the exact sizes of the gaps are unknown. This record will be updated with the finished sequence as soon as it is available and the accession number will be preserved.

1 852: contig of 852 bp in length

*

* 853	872: gap of unknown length	* 33181	34033: contig of 853 bp in length
* 873	1753: contig of 881 bp in length	* 34034	gap of unknown length
* 1754	1773: gap of unknown length	* 34054	contig of 919 bp in length
* 1774	2658: contig of 885 bp in length	* 34973	gap of unknown length
* 2659	2678: gap of unknown length	* 34993	contig of 883 bp in length
* 2679	3514: contig of 836 bp in length	* 35876	gap of unknown length
* 3515	3534: gap of unknown length	* 35896	contig of 836 bp in length
* 3535	4449: contig of 915 bp in length	* 36732	gap of unknown length
* 4450	4469: gap of unknown length	* 36752	contig of 888 bp in length
* 4470	5384: contig of 915 bp in length	* 37640	gap of unknown length
* 5385	5404: gap of unknown length	* 37660	contig of 840 bp in length
* 5405	6273: contig of 869 bp in length	* 38500	gap of unknown length
* 6274	6293: gap of unknown length	* 38520	contig of 829 bp in length
* 6294	7095: contig of 802 bp in length	* 39349	gap of unknown length
* 7096	7115: gap of unknown length	* 39369	contig of 1009 bp in length
* 7116	7981: contig of 866 bp in length	* 40378	gap of unknown length
* 7982	8001: gap of unknown length	* 40398	contig of 1072 bp in length
* 8002	8729: contig of 728 bp in length	* 41469	gap of unknown length
* 8730	8749: gap of unknown length	* 41470	contig of 401 bp in length
* 8750	9875: contig of 1126 bp in length	* 41490	gap of unknown length
* 9876	9895: gap of unknown length	* 41891	contig of 885 bp in length
* 9896	10631: contig of 736 bp in length	* 42796	gap of unknown length
* 10632	10651: gap of unknown length	* 42816	contig of 929 bp in length
* 10652	11577: contig of 926 bp in length	* 43745	gap of unknown length
* 11578	11597: gap of unknown length	* 43765	contig of 864 bp in length
* 11598	12454: contig of 857 bp in length	* 44628	gap of unknown length
* 12455	12474: gap of unknown length	* 44629	contig of 1310 bp in length
* 12475	13334: contig of 860 bp in length	* 44649	gap of unknown length
* 13335	13354: gap of unknown length	* 45959	contig of 849 bp in length
* 13355	14279: contig of 925 bp in length	* 45979	gap of unknown length
* 14280	14299: gap of unknown length	* 46828	contig of 960 bp in length
* 15245	15245: contig of 946 bp in length	* 46848	gap of unknown length
* 15246	15265: gap of unknown length	* 47808	contig of 549 bp in length
* 15266	16058: contig of 793 bp in length	* 47828	gap of unknown length
* 16059	16078: gap of unknown length	* 48377	contig of 901 bp in length
* 16079	16943: contig of 867 bp in length	* 48397	gap of unknown length
* 16946	16965: gap of unknown length	* 49298	contig of 865 bp in length
* 16966	17850: contig of 885 bp in length	* 49318	gap of unknown length
* 17851	17870: gap of unknown length	* 50183	contig of 852 bp in length
* 17871	18722: contig of 852 bp in length	* 50203	gap of unknown length
* 18723	18743: gap of unknown length	* 51055	contig of 299 bp in length
* 18743	19550: contig of 814 bp in length	* 51075	gap of unknown length
* 19557	19578: gap of unknown length	* 51374	contig of 660 bp in length
* 19577	20816: contig of 1240 bp in length	* 51394	gap of unknown length
* 20837	20836: gap of unknown length	* 52054	contig of 871 bp in length
* 21757	21756: contig of 920 bp in length	* 52944	gap of unknown length
* 21777	21775: gap of unknown length	* 52964	contig of 575 bp in length
* 22518	22517: contig of 741 bp in length	* 53539	gap of unknown length
* 22518	22531: gap of unknown length	* 53559	contig of 815 bp in length
* 22538	23457: contig of 920 bp in length	* 54374	gap of unknown length
* 23458	23477: gap of unknown length	* 54394	contig of 859 bp in length
* 23478	24289: contig of 812 bp in length	* 55253	gap of unknown length
* 24290	24309: gap of unknown length	* 55273	contig of 1254 bp in length
* 24310	25249: contig of 940 bp in length	* 55274	gap of unknown length
* 25250	25260: gap of unknown length	* 56527	contig of 829 bp in length
* 25270	26133: contig of 866 bp in length	* 56547	gap of unknown length
* 26136	26153: gap of unknown length	* 57377	contig of 77 bp in length
* 26156	27048: contig of 893 bp in length	* 57396	gap of unknown length
* 27049	27068: gap of unknown length	* 57473	contig of 844 bp in length
* 28003	28003: contig of 935 bp in length	* 58337	gap of unknown length
* 28004	28023: gap of unknown length	* 58357	contig of 87 bp in length
* 28024	28874: contig of 851 bp in length	* 58444	gap of unknown length
* 28875	28894: gap of unknown length	* 58464	contig of 824 bp in length
* 28895	29777: contig of 883 bp in length	* 59288	gap of unknown length
* 29778	29797: gap of unknown length	* 59308	contig of 1191 bp in length
* 29798	30160: contig of 363 bp in length	* 60499	gap of unknown length
* 30161	30180: gap of unknown length	* 60519	contig of 1340 bp in length
* 30181	31014: contig of 834 bp in length	* 61859	gap of unknown length
* 31015	31034: gap of unknown length	* 61879	contig of 873 bp in length
* 31035	31880: contig of 846 bp in length	* 62752	gap of unknown length
* 31881	31900: gap of unknown length	* 62772	contig of 851 bp in length
* 31901	33160: contig of 1260 bp in length	* 63623	gap of unknown length
* 33161	33180: gap of unknown length	* 64694	contig of 1051 bp in length

* arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlapping relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

* 43205 44325: contig of 1121 bp in length
* 44326 44425: gap of unknown length
* 44426 45890: contig of 1465 bp in length
* 45891 45990: gap of unknown length
* 45991 47229: contig of 1239 bp in length
* 47230 47329: gap of unknown length
* 47330 48645: contig of 1316 bp in length
* 48646 48745: gap of unknown length
* 48746 49980: contig of 1235 bp in length
* 49981 50080: gap of unknown length
* 50081 51209: contig of 1129 bp in length
* 51210 51309: gap of unknown length
* 51310 52798: contig of 1489 bp in length
* 52799 52898: gap of unknown length
* 52899 54154: contig of 1256 bp in length
* 54155 54254: gap of unknown length
* 54255 55552: contig of 1298 bp in length
* 55553 55652: gap of unknown length
* 55653 57780: contig of 2128 bp in length
* 57781 57880: gap of unknown length
* 57881 59692: contig of 1812 bp in length
* 59693 59792: gap of unknown length
* 59793 60947: contig of 1155 bp in length
* 60948 61047: gap of unknown length
* 61048 62132: contig of 1085 bp in length
* 62133 62232: gap of unknown length
* 62233 63505: contig of 1273 bp in length
* 63506 63605: gap of unknown length
* 63606 65114: contig of 1509 bp in length
* 65115 65214: gap of unknown length
* 65215 67705: contig of 2491 bp in length
* 67706 67805: gap of unknown length
* 67806 69266: contig of 1461 bp in length
* 69267 69366: gap of unknown length
* 69367 70408: contig of 1042 bp in length
* 70409 70509: gap of unknown length
* 70509 72169: contig of 1661 bp in length
* 72170 72269: gap of unknown length
* 72270 73711: contig of 1442 bp in length
* 73712 73811: gap of unknown length
* 73812 76440: contig of 2629 bp in length
* 76441 76540: gap of unknown length
* 76541 77916: contig of 1376 bp in length
* 77917 78016: gap of unknown length
* 78017 80190: contig of 2174 bp in length
* 80191 80290: gap of unknown length
* 80291 81379: contig of 1089 bp in length
* 81380 81479: gap of unknown length
* 81480 82828: contig of 1349 bp in length
* 82829 82928: gap of unknown length
* 82929 86817: contig of 3889 bp in length
* 86818 86917: gap of unknown length
* 86919 89152: contig of 2235 bp in length
* 89153 89252: gap of unknown length
* 89253 91565: contig of 2313 bp in length
* 91566 91665: gap of unknown length
* 91666 94073: contig of 2408 bp in length
* 94074 94174: gap of unknown length
* 94174 98664: contig of 4491 bp in length
* 98665 98764: gap of unknown length
* 98765 101607: contig of 2843 bp in length
* 101608 101707: gap of unknown length
* 101708 103901: contig of 2194 bp in length
* 103902 104001: gap of unknown length
* 104002 105900: contig of 1899 bp in length
* 105901 106000: gap of unknown length
* 106001 109702: contig of 3702 bp in length
* 109703 109802: gap of unknown length
* 109803 116164: contig of 6362 bp in length
* 116165 116264: gap of unknown length
* 116265 126230: contig of 9966 bp in length
* 126231 126330: gap of unknown length
* 126331 145648: contig of 19318 bp in length

FEATURES
source
1..298166
Location/Qualifiers
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="16"
/clone="RP11-358F6"
/clone_lib="RP11 human BAC library 11"

Query Match 21.5%; Score 132.8; DB 78; Length 298166;
Best Local Similarity 60.9%; Pred. No. 5.6e-28;
Matches 266; Conservative 0; Mismatches 168; Indels 3; Gaps 3;

QY 4 ttgatgttattctcacccgtcgccatggtgcacattgtggcgttatgagcccggtccc 63
Db 90711 TTAGAGGTATTTTACACTAANAATCGTCACATTGTGGCGCTGATGA-CCAAGGGCCC 90769

QY 64 gattcttttttgcctctcagacacgctgcagtcgtcccgtaaaagagcgatgatgggc 123
Db 90770 TATTCTTTATGGGATGTGAGACCGGAGTCAATCTTCGCGAAGAGAATCTACGATGGG 90829

QY 124 gtgctgggcattaccctcgccggttaattgttggcgtggattgcgtgcttggcctgc 183
Db 90830 GAGGTGGCCATTACCTCGTCTGATGTTGGCGTGGATAGCTCTGGCTCAGCGCAT 90889

QY 184 ttgattatcgaaaaaattggcctgctgcatacgtgattatgttggcgtgcttgcctgtat 243
Db 90890 GGGG-TATTCAAAGAAATGATCTGTGTACTCTCTAGCCGGC-GTTGCCCTGGAA 90947

QY 244 ctctgctggatgggttaccagatgctacgtgtgctacgtgaaagagggcggtttctgca 303
Db 90948 CGATGTTGAAAGAGTTAAAGATGCTTGTCGCGCGGTGATGGAGGAGACGTCTCCTTC 91007

QY 304 cctgcgcacaggtgcagctggcgaaaagtggcgagattcctctgaaagtttactgacc 363
Db 91008 TCTGTGCTAGGGGGAGGTGCTTAACTAGGGGGGCAACTCGGTGAGAGGGCCACTGACC 91067

QY 364 aatctcgctaaccgaaagcattatctactttgtgctcggtgttcttcttcttcttctggt 423
Db 91068 AATCACCCTCCCTGAAACTATTAAAGCCGTAGAGTGGCGATCCGAACTCGGTGGGT 91127

QY 424 gataacgttggcactac 440
Db 91128 GTTAATGGGAATCCAC 91144

RESULT 13
AC022157 AC022157 4833 bp DNA HTG 26-JAN-2000
LOCUS Homo sapiens chromosome 5 clone RP1-280K18, LOW-PASS SEQUENCE
DEFINITION SAMPLING.
ACCESSION AC022157
VERSION AC022157.1 GI:6758618
KEYWORDS HTG; HTGS_PHASE0.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 4833)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 5
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 4833)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (26-JAN-2000) Production Sequencing Facility, DOE Joint

COMMENT	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
	-----Genome Center	
JOURNAL REFERENCE	Center: Joint Genome Institute	
	Center Code: JGI	
AUTHORS	Web site: http://www.jgi.doe.gov	

TITLE	* NOTE: This record contains 8 individual	
	* sequencing reads that have not been assembled into	
JOURNAL	* contigs. Runs of N are used to separate the reads	
	* and the order in which they appear is completely	
COMMENT	* arbitrary. Low-pass sequence sampling is useful for	
	* identifying clones that may be gene-rich and allows	
	* overlap relationships among clones to be deduced.	
	* However, it should not be assumed that this clone	
	* will be sequenced to completion. In the event that	
	* the record is updated, the accession number will	
	* be preserved.	
	* 1	
	* 836: contig of 836 bp in length	
	* gap of unknown length	
	* 837	
	* 1566: contig of 730 bp in length	
	* gap of unknown length	
	* 1567	
	* 2612: contig of 1046 bp in length	
	* gap of unknown length	
	* 2613	
	* 2721: contig of 109 bp in length	
	* gap of unknown length	
	* 2722	
	* 2829: contig of 108 bp in length	
	* gap of unknown length	
	* 2830	
	* 3052: contig of 223 bp in length	
	* gap of unknown length	
	* 3053	
	* 4256: contig of 1204 bp in length	
	* gap of unknown length	
	* 4257	
	* 4833: contig of 577 bp in length.	
FEATURES	Location/Qualifiers	
	1. .4833	
source	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/chromosome="5"	
	/clone="RP1-280K18"	
BASE COUNT	1218 a	1298 c 1250 g 1055 t 12 others
	ORIGIN	
Query Match	19.4%;	Score 120; DB 67; Length 4833;
	Best Local Similarity	91.4%; Pred. No. 2.4e-24;
Matches 139; Conservative 0; Mismatches 10; Indels 3; Gaps 1;		
Qy 449	ggggcatttttcgctgatcattgtcgaaacgctggcggtgttaacctgcg---ttgccca 505	
Db 25	GCGCGCATTTTATAGCTGATCATTTGCGAAACGCTGGCGTGTATACCGCTCGTTGCCAC 84	
Qy 506	gcctgtttgcccgcgcaaatgcgcgtggttatcaacgtctgcggaagtgtgattgatg 565	
Db 85	GCCTGTTTGCCTCGCGCAAAATGCGCGCGTGTATCAACGCTCGCGGAAGTGATTGATG 144	
Qy 566	gtttgcggggcggttatttgcggatttggc 597	
Db 145	GTTTTCGCGGGCGGCTATTATTGCGGATTTGGC 176	
RESULT 14		
AC020874/c	AC020874	265383 bp DNA HTG 16-FEB-2000
LOCUS	Mus musculus clone RP23-333M14, LOW-PASS SEQUENCE SAMPLING.	
DEFINITION	AC020874	
ACCESSION	AC020874.2	
VERSION	GI:6980207	
KEYWORDS	HTG; HTGS_PHASE0.	
SOURCE	house mouse.	
ORGANISM	Mus musculus	
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;	
	Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.	
AUTHORS	1 (bases 1 to 265383)	
	DOE Joint Genome Institute.	
TITLE	Sequencing of Mouse	

Unpublished	2 (bases 1 to 265383)	
	DOE Joint Genome Institute.	
JOURNAL REFERENCE	Direct Submission	
	Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint	
AUTHORS	Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA	
	On Feb 16, 2000 this sequence version replaced gi:6686434.	
TITLE	* NOTE: This record contains 244 individual	
	* sequencing reads that have not been assembled into	
JOURNAL	* contigs. Runs of N are used to separate the reads	
	* and the order in which they appear is completely	
COMMENT	* arbitrary. Low-pass sequence sampling is useful for	
	* identifying clones that may be gene-rich and allows	
	* overlap relationships among clones to be deduced.	
	* However, it should not be assumed that this clone	
	* will be sequenced to completion. In the event that	
	* the record is updated, the accession number will	
	* be preserved.	
	* 1	
	* 869: contig of 869 bp in length	
	* gap of unknown length	
	* 870	
	* 1181: contig of 312 bp in length	
	* gap of unknown length	
	* 1182	
	* 2202: contig of 1021 bp in length	
	* gap of unknown length	
	* 2203	
	* 2960: contig of 758 bp in length	
	* gap of unknown length	
	* 2961	
	* 3646: contig of 686 bp in length	
	* gap of unknown length	
	* 3647	
	* 4520: contig of 874 bp in length	
	* gap of unknown length	
	* 4521	
	* 4813: contig of 293 bp in length	
	* gap of unknown length	
	* 4814	
	* 5434: contig of 621 bp in length	
	* gap of unknown length	
	* 5435	
	* 6000: contig of 566 bp in length	
	* gap of unknown length	
	* 6001	
	* 5992: contig of 992 bp in length	
	* gap of unknown length	
	* 6993	
	* 7785: contig of 793 bp in length	
	* gap of unknown length	
	* 7786	
	* 8367: contig of 582 bp in length	
	* gap of unknown length	
	* 8368	
	* 9114: contig of 747 bp in length	
	* gap of unknown length	
	* 9115	
	* 9729: contig of 615 bp in length	
	* gap of unknown length	
	* 9730	
	* 10313: contig of 584 bp in length	
	* gap of unknown length	
	* 10314	
	* 10712: contig of 399 bp in length	
	* gap of unknown length	
	* 10713	
	* 11404: contig of 692 bp in length	
	* gap of unknown length	
	* 11405	
	* 12261: contig of 857 bp in length	
	* gap of unknown length	
	* 12262	
	* 13173: contig of 912 bp in length	
	* gap of unknown length	
	* 13174	
	* 13934: contig of 761 bp in length	
	* gap of unknown length	
	* 13935	
	* 14703: contig of 769 bp in length	
	* gap of unknown length	
	* 14704	
	* 15380: contig of 677 bp in length	
	* gap of unknown length	
	* 15381	
	* 15654: contig of 274 bp in length	
	* gap of unknown length	
	* 15655	
	* 16714: contig of 1060 bp in length	
	* gap of unknown length	
	* 16715	
	* 17249: contig of 535 bp in length	
	* gap of unknown length	
	* 17250	
	* 17863: contig of 614 bp in length	
	* gap of unknown length	
	* 17864	
	* 18508: contig of 645 bp in length	
	* gap of unknown length	
	* 18509	
	* 19193: contig of 685 bp in length	

* gap of unknown length
* 19194 19737: contig of 544 bp in length
* gap of unknown length
* 19738 20600: contig of 863 bp in length
* gap of unknown length
* 20601 21239: contig of 639 bp in length
* gap of unknown length
* 21240 21903: contig of 664 bp in length
* gap of unknown length
* 21904 22705: contig of 802 bp in length
* gap of unknown length
* 22706 22948: contig of 243 bp in length
* gap of unknown length
* 22949 23682: contig of 734 bp in length
* gap of unknown length
* 23683 24364: contig of 682 bp in length
* gap of unknown length
* 24365 25288: contig of 924 bp in length
* gap of unknown length
* 25289 25954: contig of 666 bp in length
* gap of unknown length
* 25955 26094: contig of 140 bp in length
* gap of unknown length
* 26095 27015: contig of 921 bp in length
* gap of unknown length
* 27016 27740: contig of 725 bp in length
* gap of unknown length
* 27741 28060: contig of 320 bp in length
* gap of unknown length
* 28061 28343: contig of 283 bp in length
* gap of unknown length
* 28344 29016: contig of 673 bp in length
* gap of unknown length
* 29017 29371: contig of 355 bp in length
* gap of unknown length
* 29372 29586: contig of 215 bp in length
* gap of unknown length
* 29587 29691: contig of 105 bp in length
* gap of unknown length
* 29692 30501: contig of 810 bp in length
* gap of unknown length
* 30502 31132: contig of 631 bp in length
* gap of unknown length
* 31133 31249: contig of 117 bp in length
* gap of unknown length
* 31250 32467: contig of 1218 bp in length
* gap of unknown length
* 32468 32878: contig of 411 bp in length
* gap of unknown length
* 32879 34069: contig of 1191 bp in length
* gap of unknown length
* 34070 34694: contig of 625 bp in length
* gap of unknown length
* 34695 35933: contig of 1239 bp in length
* gap of unknown length
* 35934 36920: contig of 987 bp in length
* gap of unknown length
* 36921 37173: contig of 253 bp in length
* gap of unknown length
* 37174 37282: contig of 109 bp in length
* gap of unknown length
* 37283 38196: contig of 914 bp in length
* gap of unknown length
* 38197 38452: contig of 256 bp in length
* gap of unknown length
* 38453 38678: contig of 226 bp in length
* gap of unknown length
* 38679 38835: contig of 157 bp in length
* gap of unknown length
* 38836 39561: contig of 726 bp in length
* gap of unknown length
* 39562 40452: contig of 891 bp in length
* gap of unknown length

* 40453 41223: contig of 771 bp in length
* gap of unknown length
* 41224 41864: contig of 641 bp in length
* gap of unknown length
* 41865 42690: contig of 826 bp in length
* gap of unknown length
* 42691 42988: contig of 298 bp in length
* gap of unknown length
* 42989 43612: contig of 624 bp in length
* gap of unknown length
* 43613 44265: contig of 653 bp in length
* gap of unknown length
* 44266 45263: contig of 998 bp in length
* gap of unknown length
* 45264 46592: contig of 1329 bp in length
* gap of unknown length
* 46593 47054: contig of 462 bp in length
* gap of unknown length
* 47055 47193: contig of 139 bp in length
* gap of unknown length
* 47194 48027: contig of 834 bp in length
* gap of unknown length
* 48028 48670: contig of 643 bp in length
* gap of unknown length
* 48671 49511: contig of 841 bp in length
* gap of unknown length
* 49512 49895: contig of 384 bp in length
* gap of unknown length
* 49896 50124: contig of 229 bp in length
* gap of unknown length
* 50125 50776: contig of 652 bp in length
* gap of unknown length
* 50777 50905: contig of 129 bp in length
* gap of unknown length
* 50906 51260: contig of 355 bp in length
* gap of unknown length
* 51261 52029: contig of 769 bp in length
* gap of unknown length
* 52030 52294: contig of 265 bp in length
* gap of unknown length
* 52295 53140: contig of 846 bp in length
* gap of unknown length

Query Match 14.8%; Score 91.4; DB 66; Length 265383;
Best Local Similarity 90.1%; Pred. No. 8.2e-16;
Matches 109; Conservative 0; Mismatches 11; Indels 1; Gaps 1;
QY 1 atgttgatgttatttctcacgcgtgcacattgtgcgc-ttatgagccgg 59
|||||
Db 124944 ATGTTGATGTTATTCTACCGTCGCCATTTGCGCATTTATGAGCCCG 124885
QY 60 tccgattcttttctctcagaccgtgcagtcgtccgtaaagaagcgatgat 119
|||||
Db 124884 TCCCGATTTCTTTTGTCTCTCAGACCGCTGTCAGTCGTCCTCTAAGAGATTTCT 124825
QY 120 g 120
|
Db 124824 G 124824

RESULT 15
AE006110/c
LOCUS AE006110 10184 bp DNA BCT 07-FEB-2001
DEFINITION Pasteurella multocida PM70 section 77 of 204 of the complete genome.
ACCESSION AE006110 AE004439
VERSION AE006110.1 GI:12721018
KEYWORDS Pasteurella multocida.
SOURCE Pasteurella multocida.
ORGANISM Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
REFERENCE 1 (bases 1 to 10184)

AUTHORS	May,B.J., Zhang,Q., Li,L., Paustian,M.L., Whittam,T.S. and Kapur,V.S.
TITLE	Complete nucleotide sequence of an avian isolate of Pasteurella multocida
JOURNAL	Proc. Natl. Acad. Sci. U.S.A. (2001) In press
REFERENCE	2 (bases 1 to 10184)
AUTHORS	Zhang,Q. and Kapur,V.
TITLE	Direct Submission
JOURNAL	Submitted (24-OCT-2000) Department of Veterinary Pathobiology, University of Minnesota, 1971 Commonwealth Ave., St. Paul, MN 55108, USA
FEATURES	Location/Qualifiers
source	1..10184
	/organism="Pasteurella multocida"
	/strain="PM70"
	/db_xref="taxon:747"
gene	105..1820
	/gene="ttrS"
CDS	/note="PM0724"
	105..1820
	/gene="ttrS"
	/codon_start=1
	/transl_table=11
	/product="ttrS"
	/protein_id="AAK02808.1"
	/db_xref="GI:12721019"
	/translation="MKKCSFLKMFILLSEFPALVYAQQWKIGILAQRGVAHTATW QPMWLNQHLSPHFTLVTDLNDLIQDQAGVDFILSNQAGFFYLNNNEVRWLVL RSPROQSSMGRIGSAILVRDSPFPQLDRLHKTVSAVDNAGFGFLGILYNEFYQQ GLVENKDFOLKFTGFPVDNLTLLIQERKVDAAIIVPGVILEELNEGVYKKEDFRLLAG KENVLGKASTELLPNWSLAALSQVPDSLASKVLTLLOSVPAPHLPOWTPYSTSQAD TVLGLKHPQOOLWDSKYWLOQYCGMIIFVLLLVNLVLTISFQIHRKSKALQOAY REMHYEQQLQADRLLISLGMSTGIAHEINQPLSARIMYVEGVKYHLAOLPAKNTIV EILDVQVQDRSDRIIRNLNMAKGRKTPSSCEVALKPLLRVYSFMSLQTHQKAOI NLNCPASKIKIDITLLEQVLCNLLNASQAGSEIEVNVKAGVEYMTIHDVNGSGF KAGELAPFPVFRSSKANGLGGLVLCQRLMRISINGDISLHNRLDQOQGAEVILKLPNK VGETC"
gene	1814..2392
	/gene="ttrR"
	/note="PM0725"
CDS	1814..2392
	/gene="ttrR"
	/codon_start=1
	/transl_table=11
	/product="ttrR"
	/protein_id="AAK02809.1"
	/db_xref="GI:12721020"
	/translation="MLIHLVDDDLTVLDAARFLLEQAGYQVQTSNSQAQFVNKAPLPE PGVVLDMKMPLLDGHQVHFQFLCQOQSTLAVYIMTAHADVPMAVQELKQGAVDLQKP VOFSQLQSVLKTATQTTQTSRYRYKIRHCYAQLSRKELDILALLIQGCINRQIAEALN ISVTVVHRSHIMEKQAOITAEIIVYKTAQL"
gene	complement(2427..3239)
	/gene="dapB"
	/note="PM0726"
CDS	complement(2427..3239)
	/gene="dapB"
	/codon_start=1
	/transl_table=11
	/product="dapB"
	/protein_id="AAK02810.1"
	/db_xref="GI:12721021"
	/translation="MTRVAIVGAGGRMRQLIQAVSETGVVLGAAPERQSGSLLGA DAGELANYHGLGVQITDOLASQODFDLLIDTRPEGTLAHLAFCAQKKNKVIQTTG FDDGKAKIQQAADSIGIVFASNSFVGNLVFKLEAKAVMGDYCDIEIIEAHRHK VDAFSGTALSNGEHAKTGLRDLKTHGVTRDGLTGERKRDEIGFATIRASDVGEHS VWFADIGERIEAHTKATSRWTFKAGCAVRAAKWLAQKEGLFDMTVDVLDNLN"
gene	complement(3294..3938)
	/gene="PM0727"
CDS	complement(3294..3938)
	/gene="PM0727"
	/codon_start=1
	/transl_table=11
	/product="unknown"
	/protein_id="AAK02811.1"
	/db_xref="GI:12721022"
	/translation="MNLNLIIVHFFGLITPGDFFVVSRLAASNSRNALCAVIGITL GVLFWALASILGLAILFNFTVPVLOGLWMLGGYLAJLGYLMKSOQNVFEPVSQOE QNKQTSIKKEITKGLLVNLSNAKAVIYFASVMSLVVNLITQWQINSALLIIVLETFI YFAISIVFSRQQAQOFYSQSRVIDHLSGVIFLLFGVYMLSGSIGEMTFLISK"
gene	complement(3944..4429)
	/gene="pgpA"
	/note="PM0728"
CDS	complement(3944..4429)
	/gene="pgpA"
	/codon_start=1
	/transl_table=11
	/product="PgpA"
	/protein_id="AAK02812.1"
	/db_xref="GI:12721023"
	/translation="MLKTPLOQLKLSNPTHFLALGFCAGLIRPAGTGWGSMVGLMLGW GLIQLNLGSLFFLTALCFALGCYLCKOATREDMGVHDHGAIVWDEIVGIFVLLSLPN LSPFWCLTAFILFRFFDILKPYPIRYFDHKLSESGFIMLDDIILAAIYAVLVILRLM F"
gene	complement(4439..5431)
	/gene="thil"
	/note="PM0729"
CDS	complement(4439..5431)
	/gene="thil"
	/codon_start=1
	/transl_table=11
	/product="thil"
	/protein_id="AAK02813.1"
	/db_xref="GI:12721024"
	/translation="MSHGFEDILQRYTFAKRTAKKDVILSIGDCAITELRONORLV ITDTWVTHFELPSMHAADLAYKAAVTNLSDLAAMGAEPANVSLAULTDPVDAQWLS TFGSFFDVLHDYNDLIGDGTGKPLSVITTAHTIIPKGLGCLRSASVGDWLIYISG DFGSAAAGLSLLSQOTPLNSDHEFLIQRHLRTPRVLLGLELAHSVANAADISDG FLADLDHLTRSCQGVIELDKLPSSLVNVFVGEQATFALSGDEYELCFTVPDS NKAKLERALAYIGVYTCVGQVRAISPNQKKRIKFORGNPIDITIQSGFDHFK"
gene	complement(5448..5882)
	/gene="nusb"
	/note="PM0730"
CDS	complement(5448..5882)
	/gene="nusb"
	/codon_start=1
	/transl_table=11
	/product="Nusb"
	/protein_id="AAK02814.1"
	/db_xref="GI:12721025"
	/translation="MTEQKHEKKISPRRRARECAVALYSWYVSONSPAIEIENFMAE QDLGVDTAYFRRLLPROTAENVDAVDNIMIPYLDREVSELDPIEKAILRLAYELKPE LDVPYKVVINEAIEVAKVGAEDSHKYVNGVLDKVAIVLSRK"
gene	complement(5888..6361)
	/gene="ribh"
	/note="PM0731"
CDS	complement(5888..6361)
	/gene="ribh"
	/codon_start=1
	/transl_table=11
	/product="ribh"
	/protein_id="AAK02815.1"
	/db_xref="GI:12721026"
	/translation="MKVLEGAAFPNAKIAVVIARFNSFINESLLEGALDALKRIGQV KEENTILVRVPGAVELPLVARIAESKKYDGIVALGTIVRGTAHFHYVAGEASSGLG QVANNADIPVAFGVLTETIEQAIERAGTKAGNKAGAAALVALEMVNLLAQIDAA"
gene	6575..7786
	/gene="tyrp_1"
	/note="PM0732"
CDS	6575..7786
	/gene="tyrp_1"
	/codon_start=1
	/transl_table=11
	/product="tyrp"
	/protein_id="AAK02816.1"
	/db_xref="GI:12721027"
	/translation="MKNKTFGSALLVAGTTIGAGMLAMPLTSAEMGFTYVTLILFILM

GLSYSALLFEVYOKAETKNAGIATLAEOYGLPGRILATLSLVFVYAILSAVVTG
GGSLAGVLPELGEHATPLISVSFTIVLGVFIVISGAVDMLTLEFLMKLVAFVFL
LMMLPLVNGNLLAMFLKFLIISAPVFTSFGFHVLIIPSINSILDDGNIRLRITAI
TGTAIPLVAITIIWOLATHGVFPQAFVQIINLTDPTLNGLITATYQATSAIISHA
FPTALITSLFGLVALSLFDCLLDLLKRVKIKTNRVSLGLITLPLIFALFVGEFVM
ALGYAGOMFTFYGLVLFPVGMARARKRPDLPYRVMGNNLTFLGAILILIMNVPFL
IKAGYLPFVIG"
complement(7842. .8504)
/gene="PM0733"
CDS complement(7842. .8504)
/gene="PM0733"
/codon_start=1
/transl_table=11
/product="unknown"
/db_xref="GI:12721028"
/translation="MSQHTKQOVLSAFHRSAIRRVDPNKKYSEEDFHYILELGR
SSVGSERWHLVQDRLREQLKAVSWGHTQDLDSDYLVF ILAKKHARYDSFYFHD
LAKGLFETQM KALAIYIERFQDDIOVLNERTLFDWACKQTYIALANMMTGAAMIG
IDSCPTEGFQYCAVNNQLLSQTGAFPNEYGSVAVTFGYRAKEVRPKSRKPLTEITW
LK"
8727. .10106
/gene="htra"
/note="PM0734"
8727. .10106
/gene="htra"
/codon_start=1
/transl_table=11
/product="Htra"
/protein_id="AAK02818.1"
/db_xref="GI:12721029"
/translation="MRKRRLTSLIALGSLISLTPMISOASLPQVDSGOVPSLAPML
EKVLPAVVSISVGKAKNGSTPFQGIPEEPRFFGPDLDNDRAPNFRGSGVLIIN
AERGVLTNNHVIDSADKITVQDGRKLSAKVIGTDELSVDLALIOIEKPNLTLTI
ADSKLRVGFDPFNPLGQVTSIGVISALGRSTGSDNGTYENYIQTDAAVNRGN
SGPLNLQGLIGINTAIISPSGNAGIAFAIPSNMANNVQQLIEFEVRRGMLGI
KGGELNADLAKAFDIEAQCAFVSEVLPSNAAEKAGLKAGDVIVAMNGOKISSFAEMR
AKIATSGAGKEIALTYLRDCKTHOTKVTLQSDODTOADASHLLPALAGAEMSNHDDKG
VKGVLITSVTPKSLAEQRLKKGDVIIGVNRQKVENLQQLRKILDAKPSATALNVRG
NTNFXLLIQ"

BASE COUNT 3070 a 2201 c 1937 g 2976 t
ORIGIN

Query Match 14.2%; Score 88; DB 1; Length 10184;
Best Local Similarity 48.1%; Pred. No. 6.5e-15;
Matches 281; Conservative 0; Mismatches 300; Indels 3; Gaps 1;
QY 28 atggtgcacattgtggcgcttatgagcccggtgccgattctttttgtctctcagacc 87
Db 3920 ATTGTACATTTTTCGGTTTAATCAGCCAGGACCCGACTTTTATGTCAGTCGATTA 3861
QY 88 gctgtcagctgtccgtaagagcgatgatggcgctgctggcattaccctcggcgta 147
Db 3860 GCTGCCAGTAATCAGTCGTAATGCAGCTTTGTGCCGTGATTGGTAFTACCTTAGGGGGT 3801
QY 148 atggtttgggtgggattggcgctgtggcctgattgattatcgaaaaaatggcctgg 207
Db 3800 TTGTTTGGGCATTAGCCTCTATTATTAGTTTAGCCATATTATTAAACGGTTCCTGTT 3741
QY 208 ctgcatacgcgtgattatggcggtggcgctgtatctctgctggatgggttaccagatg 267
Db 3740 TTACAGGATTAGTCATCACTTTAGCGCGGTGGGTATTTAGCCCTACTTGGGTTACCTTAG 3681
QY 268 ctacgtgtgcactgaataaagagcggtttctgcacctgcgc---cacaggtcgagctg 324
Db 3680 TTAAAAAGTCAACAAAATGTGGTTTCGAGCCAGTTTCTGAACAAGAACAAAATAAACAG 3621
QY 325 gcgaaaagtggcgagtttctcgaaggtttactgaaccaatctcgtaatccgaaagcg 384
Db 3620 ACCACATCAAAAAGAAATCACCAGAGGCGCTTTAGTGAATTTATCTAATGCTAAAGCG 3561
QY 385 attatctactttggctcgggtgtctctcattgtttgtcgggtgataacggttgacctaccgcg 444

Db 3560 GTGATTTACTTTGCTAGTGTGATGCTCTTTGGTTTTAGTTAAATTAACCCAAACATGGCAG 3501
QY 445 cgcctggggcatttttgcgcgtgacatcattgtcgaaacgcctggcggtttaccctcgttgc 504
Db 3500 ATCTGGAGCGCACATTTTGATCATTTGATATTGGAACATTTCTCTGATTATTTATGCATTCA 3441
QY 505 agcctgtttgcctgcgcgcaaatgcgcctgggtgttatcaacgctcgtgcgaagtggttatg 564
Db 3440 ATTGTTTTTTCACGTCAACAGCGCAACGACTTTTATAGTCAATATAGCCGATATATTGAT 3381
QY 565 ggttttgcggggcgcttatttgcgcggaatttggcatttcatttgc 608
Db 3380 CATCTTTCAGGCGTGATTTTCTCTGCTGTTGGTGTGTATTAAAT 3337

Search completed: May 6, 2001, 16:29:35
Job time: 4796 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 6, 2001, 15:43:39 ; Search time 132.88 Seconds

(without alignments)
2715.058 Million cell updates/sec

Title: US-09-466-935-3_COPY_187_804

Perfect score: 618

Sequence: 1 atgttgatgtattttctcac.....ttcatttgattatttcgcgg 618

Scoring table: IDENTITY_NUC

Gapop 10.0 , Gapext 1.0

Searched: 678276 seqs, 291890651 residues

Total number of hits satisfying chosen parameters: 1356552

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : N_Geneseq_0401.*

- 1: /SID52/gcgdata/geneseq/geneseq/NA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq/NA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq/NA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq/NA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq/NA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq/NA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq/NA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq/NA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq/NA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq/NA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq/NA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq/NA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq/NA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq/NA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq/NA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq/NA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq/NA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq/NA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq/NA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq/NA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq/NA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq/NA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	618	100.0	840	A48443	E. coli L-threonin
2	54.8	8.9	672	A52688	Escherichia coli y
3	36.2	5.9	8387	X84330	Stealth virus nucl
C 4	34.8	5.6	639	A52689	Escherichia coli y
C 5	34.8	5.6	10732	A10594	Gene encoding a su
C 6	33.4	5.4	1668	A61501	A. vitis hypersens
C 7	33.4	5.4	1668	A61502	A. vitis hypersens
C 8	32.4	5.2	565	C75485	Human OREF ORP1040
C 9	32	5.2	3288	T68841	Photobacterium lum
C 10	31.6	5.1	2178	C53939	Neisseria gonorrhoe
C 11	31.6	5.1	5059	X84332	Stealth virus nucl

12	31.2	5.0	903	11	Q06352
13	31	5.0	114955	20	X53491
14	30.8	5.0	1535	22	A89040
15	30.8	5.0	16885	17	T33535
C 16	30.6	5.0	4296	21	Z54326
C 17	30.6	5.0	16526	21	A81472
C 18	30.6	5.0	172325	21	F21613
C 19	30.6	5.0	349980	21	F21612
C 20	30.6	5.0	837096	21	A81489
C 21	30.4	4.9	3288	19	V29932
C 22	30.4	4.9	92934	21	A81473
C 23	30.4	4.9	349980	21	F21544
C 24	30.4	4.9	1244	20	V72577
C 25	30.2	4.9	1244	20	V72578
C 26	30.2	4.9	3131	19	V30456
C 27	30.2	4.9	534720	19	V30458
C 28	30.2	4.9	536165	19	V30459
C 29	30	4.9	2163	21	Z53940
C 30	30	4.9	50925	21	A81487
C 31	29.8	4.8	516	21	F21610
C 32	29.8	4.8	885	20	A79643
C 33	29.4	4.8	1065	21	A29323
C 34	29.4	4.8	1065	21	A29325
C 35	29.4	4.8	1195	21	A96225
C 36	29.2	4.7	1059	18	T73345
C 37	29.2	4.7	1100	21	A50040
C 38	29.2	4.7	2028	18	T73350
C 39	29.2	4.7	852	21	C43068
C 40	29	4.7	1934	17	T42302
C 41	29	4.7	38734	20	Z32020
C 42	29	4.7	38734	22	C90077
C 43	29	4.7	38734	22	C90077
C 44	28.8	4.7	1116	21	A66028
C 45	28.8	4.7	4760	20	X02056

ALIGNMENTS

RESULT	1
A48443	ID A48443 standard; DNA; 840 BP.
XX	AC A48443;
XX	DT 08-SEP-2000 (first entry)
DE	E. coli L-threonine resistance gene, rhtC.
XX	L-threonine resistance; L-threonine synthesis; rhtC;
KW	L-homoserine; L-valine; L-leucine; ds.
XX	OS Escherichia coli.
XX	Key Location/Qualifiers
FT	CDS 187..807
FT	/*tag= a
FT	/product= "RhtC"
XX	EP1013765-A1.
XX	28-JUN-2000.
XX	20-DEC-1999; 99EP-0125406.
XX	23-DEC-1998; 98RU-0123511.
XX	(AJIN) AJINOMOTO KK.
XX	Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;
XX	WPI: 2000-414602/36.
XX	P-PSDB; Y99596.

Modified carboxyl
Human adenosine A1
Mycobacterium tube
BCG deletion regio
Neisseria meningit
N. meningitidis pa
Neisseria meningit
Neisseria meningit
N. meningitidis pa
tcaA gene from the
N. meningitidis pa
Neisseria meningit
Mycobacterium tube
Mycobacterium tube
Human thiazide-sen
Rhizobium species
Rhizobium species
Neisseria meningit
N. meningitidis pa
Neisseria meningit
Pinus radiata cell
Clone if87_1 encod
Glycine max isofla
Glycine max isofla
DNA encoding lipas
DNA encoding Chlam
Lipase gene expres
Arabidopsis thalia
Maize ribosomal in
Human METH1 relate
AL021529 cDNA clon
E. coli proliferat
Glycine max protop

```
XX PT Novel Escherichia bacterium having enhanced L-threonine resistance due
XX PT to enhanced RhtC protein activity, used to produce L-threonine,
XX PT L-homoserine, L-valine and L-leucine -
XX PS Claim 9; Page 14-15; 24pp; English.
XX PS
XX CC The present sequence is the L-threonine resistance gene, rhtC, from
XX CC Escherichia coli. This sequence may be used to impart L-threonine
XX CC resistance on E. coli bacteria, which would be useful for producing
XX CC a high yield of L-threonine. L-threonine resistance means that the
XX CC bacteria will be able to grow on a minimal medium containing
XX CC L-threonine at a concentration at which the corresponding wild-type
XX CC strain would not grow. Since the transformed bacteria can grow on the
XX CC minimal medium, it can synthesise L-threonine, which accumulates. The
XX CC accumulated amino acids can then be removed from the culture medium.
XX CC The bacterium of the present invention may also be used to synthesise
XX CC L-homoserine, L-valine and L-leucine at increased levels.
XX SQ Sequence 840 BP; 165 A; 192 C; 248 G; 235 T; 0 other;

Query Match      100.0%; Score 618; DB 21; Length 840;
Best Local Similarity 100.0%; Pred. No. 4.1e-193;
Matches 618; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgttgatgtattttctacccgtcgccatggtgcacattgtggcgcttatgagcccggt 60
DB 187 atgttgatgtattttctacccgtcgccatggtgcacattgtggcgcttatgagcccggt 246
QY 61 cccgattctttttgtctctcagaccgtgtcagtcgtccctcgttaaaagacgatgatg 120
DB 247 cccgattctttttgtctctcagaccgtgtcagtcgtccctcgttaaaagacgatgatg 306
QY 121 ggcgtctgggcaattaccctgcggcgttaattgttggcggtggattgcgctgttggcctg 180
DB 307 ggcgtctgggcaattaccctgcggcgttaattgttggcggtggattgcgctgttggcctg 366
QY 181 catttgattatcgaaaaatggcctggctgcatacgcgtgattatgtggcggtgacctg 240
DB 367 catttgattatcgaaaaatggcctggctgcatacgcgtgattatgtggcggtgacctg 426
QY 241 tatctctgctgattgggttaccagatgctacgtgtgtgcactgaaagagcggtttct 300
DB 427 tatctctgctgattgggttaccagatgctacgtgtgtgcactgaaagagcggtttct 486
QY 301 gcaactgcgccacagtcagctgcgaaagtgaggcagtttctgaaaggtttactg 360
DB 487 gcaactgcgccacagtcagctgcgaaagtgaggcagtttctgaaaggtttactg 546
QY 361 accaatctcgctaatccgaaagcattatctactttggctggcggtgtctcattgtttgtc 420
DB 547 accaatctcgctaatccgaaagcattatctactttggctggcggtgtctcattgtttgtc 606
QY 421 ggtgataacgttggcaactaccgcgcgtgggcaatttttgcgtgatcattgtcgaacg 480
DB 607 ggtgataacgttggcaactaccgcgcgtgggcaatttttgcgtgatcattgtcgaacg 666
QY 481 ctggcggtgtttacogtctgttccagcctgtttgcctcgcgcaaatgcgcgtgtggttat 540
DB 667 ctggcggtgtttacogtctgttccagcctgtttgcctcgcgcaaatgcgcgtgtggttat 726
QY 541 caacgtctgcggaagtggattgatgttttgcggggcggttatttgcgggatttggcatt 600
DB 727 caacgtctgcggaagtggattgatgttttgcggggcggttatttgcgggatttggcatt 786
QY 601 catttgattatttcgcgg 618
DB 787 catttgattatttcgcgg 804

RESULT 2
A52688
```

```
ID XX A52688 standard; DNA; 672 BP.
AC XX A52688;
DT 03-JAN-2001 (first entry)
DE XX Escherichia coli yahN gene.
KW XX E. coli; yahN gene; amino acid production; excretion protein gene;
KW XX amino acid excretion protein; ds.
OS XX Escherichia coli.
XX FH Key Location/Qualifiers
FT CDS 1..672
FT /*tag= a
FT /product= "yahN"
XX EP1016710-A2.
XX PD 05-JUL-2000.
XX PF 17-DEC-1999; 99EP-0125263.
XX PR 30-DEC-1998; 98RU-0124016.
XX PR 09-MAR-1999; 99RU-0104431.
XX PA (AJIN ) AJINOMOTO CO INC.
XX PI Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV,
PI Tokhmakova IL;
XX DR WPI: 2000-414802/36.
XX DR P-PSDB; B01786.
XX PT Increased production of L-amino acids by an Escherichia bacterium
XX PT comprises increasing the expression amount of an L-amino acid excretion
XX PT protein -
XX PS Disclosure; Page 17-18; 29pp; English.
XX CC The present sequence is the yahN gene (an excretion protein gene) of
XX CC Escherichia coli. The amino acid excretion protein produced from this
XX CC gene is involved in the production of amino acids, and an increase in its
XX CC expression leads to an increased accumulation of amino acids in the cell.
XX CC In this case, an increase in lysine, glutamic acid and proline is
XX CC achieved if multiple copies of the gene are transfected into a bacterium.
XX CC The bacterium used is E. coli.
XX SQ Sequence 672 BP; 130 A; 154 C; 178 G; 210 T; 0 other;

Query Match      8.9%; Score 54.8; DB 21; Length 672;
Best Local Similarity 45.8%; Pred. No. 4.5e-08;
Matches 265; Conservative 0; Mismatches 307; Indels 6; Gaps 2;

QY 32 tgcacattgtggcgcttatgagcccggtcccgattctttttgtctctcagaccgctg 91
DB 77 tgttcggtattacttttttaataccggggagccaatctcttggtagtaacaaccagcc 136
QY 92 tcaatcgtcccgtaaaagaacgatgagcgctgctggcgattaccctgcggcgtaatgg 151
DB 137 tggcttcgcgtgcgcgcgcaggggtgctgacccggcggtggcgctggcgatgcat 196
QY 152 ttgggctgggattgcgctgcttggcgctgattgattatcgaaaaaatggcgctgctgc 211
DB 197 ttatttcgggttgggtttgttggcttctcaacgctaattacgcagtgtagagagattt 256
QY 212 atacgctgattatggtggcggtgacctgatctctctggtgaggttaccagatgctac 271
DB 257 ttctgcttatcagaatcgtgcggcgcttatctctt---atggttgcgtggtgcagc 312
QY 272 gtggtgcactgaaaaaagaggcggttttctcactgcgcacacaggtcgcagctggcgaaa 331
```

Db 313 atgcccgcagtgcaaacaccgcaaatagacacactacaacacgattagc--gccccc 370
QY 332 gtggcgcagtttctgaaagtgttactgacaaatctcgtactcgaagcagattatct 391
Db 371 ggtatgtctttttcgccgagattactaccgatctctcaccgcgaacacggtttat 430
QY 392 actttggcgcgtgtctctcattgttttcggtgataacgttgccactacgcgcgtggg 451
Db 431 tttttcagattttcttcagtaacataatgcgaaacacacacggtttaa 490
QY 452 gcaattttgcgtgatcatgtgcgaacgcgtggcgtgttttaacgtgttcgacgcgtg 511
Db 491 tggcctggcgggagtgctgcgcatacaattatctggcgagttttcttagcagggcgt 550
QY 512 ttgcctgcgcgaatgcgcgtgtgtttatcaacgtctgcgaagtgagtggttttg 571
Db 551 ttctttgcccgcgtgcgtcgtgtctgtatggcgatgcgaacgcgttgccagtcgggta 610
QY 572 ccgggcgctatttgcgcgatttggcattcatttgatt 609
Db 611 ttggtgcaatttgggtattcgcgtacgcgtgatt 648

RESULT 3
X84330/c
ID X84330 standard; DNA; 8387 BP.
XX AC X84330;
XX DT 08-SEP-1999 (first entry)
XX DE Stealth virus nucleic acid clone, SEQ ID NO: 22.
XX KW Stealth virus; detection; diagnosis; infection; ss.
XX OS Stealth virus.
XX FH Key Location/Qualifiers
FT misc_difference 8118 /*tag= a
FT /*note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the
FT base numbering given in the specification"
FT misc_feature 8130..8133 /*tag= b
FT /*note= "these nucleotides are represented as * in the
FT specification, and are included to maintain the
FT base numbering given in the specification"
FT misc_difference 8157 /*tag= c
FT /*note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the
FT base numbering given in the specification"
FT misc_difference 8275 /*tag= d
FT /*note= "this nucleotide is represented as a * in the
FT specification, and is included to maintain the
FT base numbering given in the specification"

PN W09334019-A1.
XX
XX PD 08-JUL-1999.
XX PF 30-DEC-1998; 98WO-US27744.
XX PR 30-DEC-1997; 97US-0001184.
XX PA (MART/) MARTIN W J.
XX PI Martin WJ;
XX WPI; 1999-405521/34.

XX Novel strains of stealth virus
XX PS Claim 19; Page 61-64; 95pp; English.
XX CC This sequence represents a stealth virus nucleic acid clone. The
CC invention relates to a method of detecting and characterising a stealth
CC virus by reacting a sample suspected of containing a stealth virus with a
CC probe from a known stealth virus and sequencing the resultant isolated
CC nucleotide. The method comprises the steps of: (a) isolating DNA or RNA
CC from a sample suspected of containing a stealth virus, e.g. a culture of
CC cells showing a viral cytopathic effect; (b) testing the reactivity of
CC the isolated DNA or RNA with a molecular probe that contains at least 18
CC or more contiguous nucleotides identical to sequence previously
CC identified from a stealth virus; and, optionally (c) sequencing the
CC isolated DNA or RNA molecules that react with the probe. The method is
CC used to detect stealth virus in a biological product, food or in the
CC environment. The method is also used to evaluate agents for their
CC inhibitory or stimulatory effects on stealth virus replication and to
CC determine capacity of the virus to recombine with and potentially alter
CC the nucleic acid sequences of a cell or bacterium.
XX SQ Sequence 8387 BP; 1807 A; 2320 C; 2218 G; 1844 T; 198 other;

Query Match 5.9%; Score 36.2; DB 20; Length 8387;
Best Local Similarity 23.5%; Pred. No. 0.2;
Matches 77; Conservative 87; Mismatches 163; Indels 0; Gaps 0;
QY 150 ggtttggcgtgattgcgtctgtgcctgcattgattatcgaaaaaatgcctgct 209
Db 8346 SNKHTSYAAMVGCWNATKYNRYTKKSANKGHNATKYNRYTKKSANKGHYVWTASTRYA 8287
QY 210 gcatacgcgtgattatggcggcgtggcctgtatctctcgtgattggttaccagatgt 269
Db 8286 NAYNKMdHGVNTTHVGNTHCHAAAGVHVHGGDHHYWASYGKNAYTKNDWTKNYNNATKND 8227
QY 270 acgtgggtgcactgaaagagagcggtttctgcacgcgccacaggtgcagtcgcgaa 329
Db 8226 WGVNAHDNATKNDWNTKNDWGVDTYVYKRSNTKYNSKACVBKGGHGVDTASNAT 8167
QY 330 aagtggcgcagttctcctgaaagtttactgacaaatctcgtactatccgaaagcattat 389
Db 8166 HNAYTKNDWNNANDTYCTYTATHTDKTYNTVGHNNNNCDMKNTYSABWRNTDWAATTCGT 8107
QY 390 ctactttgcgcgtgtctcattgtttgctgcgtgataacgttgccactacgcgcgtg 449
Db 8106 CGATCTGCTCGAGAGTCCCGCTGATTCTTCGCTGATGTGAGAAATAATGCCGCGTTC 8047
QY 450 gggcatttttgcgcgtgatcattgtcga 476
Db 8046 CGCAGCTAATCCGAGTTCACAGSCGA 8020

RESULT 4
A52689
ID A52689 standard; DNA; 639 BP.
XX AC A52689;
XX DT 03-JAN-2001 (first entry)
XX DE Escherichia coli yeas gene.
XX KW E. coli; yeas gene; amino acid production; excretion protein gene;
XX amino acid excretion protein; ds.
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
FT CDS 1..639 /*tag= a
FT FT /product= "yeas"

DT	08-FEB-2001	(first entry)	
XX	Human ORFX ORF1040	polynucleotide sequence SEQ ID NO:2079.	
DE	Human; open reading frame; ORFX; detection; cytostatic; hepatotropic; vulnery;	antipsoriatic; antiparkinsonian; nootropic; neuroprotective; immunosuppressant; cardiatic;	
XX	anticonvulsant; osteopathic; antiarthritic; vasotropic; antidiabetic; immunostimulant; thrombolytic; coagulant; immunosuppressive; antiinflammatory;		
KW	hypotensive; dermatological; immunosuppressive; antirheumatic; antithyroid;		
KW	antiviral; antibacterial; antifungal; antirheumatic; antithyroid;		
KW	antianemic; gene therapy; cancer; proliferative disorder; hypertension;		
KW	neurodegenerative disorder; osteoarthritis; graft vs host disease;		
KW	cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;		
KW	cholesterol ester storage; systemic lupus erythematosus; infection;		
KW	severe combined immunodeficiency; malaria; autoimmune disorder; asthma;		
KW	allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;		
KW	bone damage; cartilage damage; antinflammatory disease; coagulation;		
KW	thrombosis; contraceptive; ss.		
XX	Homo sapiens.		
XX	WO200058473-A2.		
XX	05-OCT-2000.		
XX	31-MAR-2000;	2000WO-US08621.	
XX	31-MAR-1999;	99US-0127607.	
PR	02-APR-1999;	99US-0127636.	
PR	03-APR-1999;	99US-0127728.	
PR	30-MAR-2000;	2000US-0540763.	
XX	(CURA-) CURAGEN CORP.		
PA	Shimkets RA, Leach M;		
XX	WPI; 2000-602362/57.		
DR	P-PSDB; B41276.		
XX	Novel nucleic acids and peptides derived from open reading frame X,		
XX	useful for treating e.g. cancers, proliferative disorders,		
PT	neurodegenerative disorders and cardiovascular disease -		
PT	Claim 5; Page 1566; 5507pp; English.		
XX	C74446 to C77606 encode the proteins given in B40237 to B43397, which		
XX	represent the human ORFX open reading frames 1 to 3161. The ORFX		
CC	sequences have activities such as: cytostatic; hepatotropic; vulnery;		
CC	antipsoriatic; antiparkinsonian; nootropic; neuroprotective; osteopathic;		
CC	anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;		
CC	cardiant; thrombolytic; coagulant; vasotropic; antidiabetic; hypotensive;		
CC	dermatological; immunosuppressive; antiinflammatory; antibacterial;		
CC	antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The		
CC	sequences can be used for determining the presence of or predisposition		
CC	to, or preventing or treating pathological conditions associated with an		
CC	ORFX-associated disorder. The nucleic acids can be used to express ORFX		
CC	proteins in gene therapy vectors. The proteins and nucleic acids may be		
CC	used to treat cancers, proliferative disorders, neurodegenerative		
CC	disorders, osteoarthritis, graft vs host disease, cardiovascular disease,		
CC	diabetes mellitus, hypertension, hypothyroidism, cholesterol ester		
CC	storage, systemic lupus erythematosus, severe combined immunodeficiency		
CC	(SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune		
CC	disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and		
CC	cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to		
CC	enhance coagulation; to inhibit thrombosis; and as a contraceptive.		
XX	Sequence 565 BP; 122 A; 169 C; 148 G; 126 T; 0 other;		
XX	Query Match	5.2%; Score 32.4; DB 21; Length 565;	
XX	Best Local Similarity	50.6%; Pred. No. 0.95;	
XX	Matches 78; Conservative	0; Mismatches 76; Indels 0; Gaps 0;	

QY	313	caggtcgagctggcgaaagtgggcgagtttctctgaaaggtttactgaccaatctcgt	372
DB	156	CAGATCGAATTCGAGGATTGCCAGTGCCCTCTCAACGCGCTTGAGAGTGATCTCGTT	97
QY	373	aatccgaaagcgattatctactttgttgctggtgttctctcattgttctcggtgataacgtt	432
DB	96	GAAGGTAATCGGTTTGTAGGATTTGACCGCCAGTACCTGCTGGATGTGCCAGCTGATGCT	37
QY	433	ggcaactaccgcgcgcgtgggcatttttgcctga	466
DB	36	TTCACTTCACGATCAGGGTTCGGTTCGAGGTAA	3
RESULT	9		
T68841/c	ID	T68841 standard; DNA; 3288 BP.	
XX	AC	T68841;	
XX	DT	29-JAN-1998 (first entry)	
XX	DE	Photorhabdus luminescens insect toxin protein Tcaa.	
XX	KW	Insecticide; insect; toxin; pest control; biological control;	
KW	Photorhabdus luminescens; Tcaa; Southern corn rootworm;		
KW	Colorado potato beetle; Western corn rootworm; meal worm;		
KW	boll weevil; turf grub; Coleoptera; beet armyworm; black cutworm;		
KW	cabbage looper; codling moth; corn earworm; European corn borer;		
KW	tobacco hornworm; tobacco budworm; Lepidoptera; Hymenoptera;		
KW	Diptera, Dictyoptera; Acarina; Homoptera; ds.		
OS	Photorhabdus luminescens.		
XX	WO9717432-A1.		
XX	15-MAY-1997.		
XX	06-NOV-1996;	96WO-US18003.	
PR	28-AUG-1996;	96US-0705484.	
PR	06-NOV-1995;	95US-0007255.	
PR	28-FEB-1996;	96US-0608423.	
XX	(WISC) WISCONSIN ALUMNI RES FOUND.		
PA	Blackburn MB, Bowen DJ, Ciche TA, Ensign JC, Fatig R;		
PI	French-Constant RH, Guo L, Hey TD, Merlo D7, Orr GL;		
PI	Petell J, Roberts JL, Rocheleau TA, Schoonover S;		
PI	Strickland JA;		
XX	WPI; 1997-281022/25.		
DR	P-PSDB; W17888-89.		
XX	Photorhabdus sp. insecticidal protein toxins and DNA encoding them -		
PT	insect control		
PT	Claim 38; Page 169-173; 276pp; English.		
PS	This genomic DNA sequence encodes insecticidal toxin protein Tcaa		
CC	(see W17888-89) of Photorhabdus luminescens, a component of		
CC	a toxin protein complex. Claimed toxins of P. luminescens (see		
CC	W17871, W17884-89, W17899-900, W18301-06) can be produced by		
CC	recombinant DNA methods and applied to, or genetically engineered		
CC	into, insect larvae food and plants for insect control. The		
CC	Photorhabdus toxins are particularly effective against Southern		
CC	corn rootworm, Colorado potato beetle, Western corn rootworm, meal		
CC	worm, boll weevil and turf grub (Coleoptera), beet armyworm, black		
CC	cutworm, cabbage looper, codling moth, corn earworm, European corn		
CC	borer, tobacco hornworm and tobacco budworm (Lepidoptera), and are		
CC	also active against insects of the orders Hymenoptera, Diptera,		
CC	Dictyoptera, Acarina and Homoptera. (All claimed).		

SQ Sequence 3288 BP; 894 A; 793 C; 768 G; 833 T; 0 other;

Query Match 5.2%; Score 32; DB 18; Length 3288;
Best Local Similarity 46.4%; Pred. No. 3;
Matches 104; Conservative 0; Mismatches 120; Indels 0; Gaps 0;

QY 159 tgggattggcgtctggcgctgcatttgaattatcagaaaaaatggcgtgctgcatacgct 218
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2728 TGGCATTGTCATGGCTAAACAGGGTTTGAATTATCATGCCAGTAGTTTGCCAAATGTAAAAG 2669

QY 219 gattatgggtggcggtggcgctgatctctgctggtggtaccagatgctacgtgtgtgc 278
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2668 GATTATCACTACCAGATTGTGATTTTCTCCAGCCAGAAGTGACGATACCCATCACATTGC 2609

QY 279 actgaaaaaaggcggtttctgcacctgcgccacagtgcagctgcycaaaagtggcg 338
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2608 TCYTATATCGGAAACCGCGCTTAGCGCCGCGACACTTCTGCTGTAGACGCCGAATCCA 2549

QY 339 cagdtcttcgaaaggtttactgaccaatctcgctaataccgaaaag 382
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 2548 TTGTTTCTTTAGTGGCAGCTTGACTATTACGCTGTCAACAACG 2505

RESULT 10
Z53939
ID Z53939 standard; DNA; 2178 BP.
XX
AC Z53939;
XX
DT 21-MAR-2000 (first entry)
XX
DE Neisseria gonorrhoeae ORF 589 partial DNA sequence SEQ ID NO:1827.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW antibiotic; diagnosis; immunogenic; infection; meningitis; septicaemia;
KW antiserum; gene therapy; ds.
XX
OS Neisseria gonorrhoeae.
XX
FN WO9957280-A2.
PD 11-NOV-1999.
XX
PF 30-APR-1999; 99WU-US093346.
XX
PR 01-MAY-1998; 98US-0083758.
PR 31-JUL-1998; 98US-0094869.
PR 02-SEP-1998; 98US-0098994.
PR 02-SEP-1998; 98US-0099062.
PR 09-OCT-1998; 98US-0103749.
PR 09-OCT-1998; 98US-0103794.
PR 09-OCT-1998; 98US-0103796.
PR 25-FEB-1999; 99US-0121528.
XX
(CHIR) CHIRON CORP.
PA (GENO-) INST GENOMIC RES.
PA
PI Fraser C, Galeotti C, Grandi G, Hickey E, Masignani V, Mora M;
PI Petersen J, Pizzo M, Rappuoli R, Ratti G, Scalato E, Scarselli M;
PI Tettelin H, Venter JC;
XX
WI WIPI; 2000-062150/05.
DR
DR
PT Novel Neisserial polypeptides predicted to be useful antigens for
PT vaccines and diagnostics -
XX
PS Claim 7; Page 918-919; 1453pp; English.


```
;
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Townsend and Townsend Khourie and Crew
; STREET: One Market Plaza, Steuart Street Tower, 20th
; FLOOR
; CITY: San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94105
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/390,878
; FILING DATE: 17-FEB-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Hunter, Tom
; REGISTRATION NUMBER: 38,498
; REFERENCE/DOCKET NUMBER: 15371A-17
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/543/9600
; TELEFAX: 415/543/5043
; INFORMATION FOR SEQ ID NO: 16:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 16885 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-08-390-878-16

Query Match 5.0%; Score 30.8; DB 1; Length 16885;
Best Local Similarity 58.9%; Pred. No. 3.8;
Matches 53; Conservative 0; Mismatches 37; Indels 0; Gaps 0;

QY 440 ccgcgcgtcgggcatttttgcgtgatcattgtcgaaacgtgcgtggtttaccgtcg 499
DB 7930 CCGCGGGGGGATCGCATTCGGGTGTTCAITGTGACGAATGCGCCCAAGCTGACCGTCG 7989
QY 500 ttgcagcgtgtttgccttcgcgcaaatgc 529
DB 7990 CGGTGCGCGGATCGCGTGCCTGCGCGGATTC 8019

RESULT 3
US-09-067-626-1/G
; Sequence 1, Application US/09067626
; Patent No. 6177086
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; APPLICANT: Nathan, Carl F.
; APPLICANT: Ehrt, Sabine
; TITLE OF INVENTION: DNA MOLECULE CONFERRING ON MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS RESISTANCE AGAINST ANTIMICROBIAL REACTIVE
; TITLE OF INVENTION: OXYGEN AND NITROGEN INTERMEDIATES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067,626
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/045,688
; FILING DATE: 06-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
```

```
;
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA: US 60/045,688
; APPLICATION NUMBER:
; FILING DATE: 06-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
; REFERENCE/DOCKET NUMBER: 19603/491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; US-09-067-626-1

Query Match 4.9%; Score 30.2; DB 4; Length 1244;
Best Local Similarity 49.7%; Pred. No. 1.4;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 185 tgattatcgaaaaaatggcctgctgcatactgctgattatgttggtggcgtgctgtatc 244
DB 595 TGCTCAGCGACGACGATCCCGTGTGCGATCCGCGGTGCGCGGCTGCGACCCCTGCATC 536
QY 245 tctgtgtagtggttaccagatctacgtgtgtgctactgaaataagagcggtttctgcac 304
DB 535 TCGCGCGCACCCGTGACGAGATGCGCGCGCCGAGGACGAGCTCGCGCGGGGACGCCACG 476
QY 305 ctgcgcacaggtgcagctgcgcaaaagtggcgcg 339
DB 475 CCGACTGCGCGATGGTGTGTCGCGCGGTGTCGCGAC 441

RESULT 4
US-09-067-626-3
; Sequence 3, Application US/09067626
; Patent No. 6177086
; GENERAL INFORMATION:
; APPLICANT: Riley, Lee W.
; APPLICANT: Nathan, Carl F.
; APPLICANT: Ehrt, Sabine
; TITLE OF INVENTION: DNA MOLECULE CONFERRING ON MYCOBACTERIUM
; TITLE OF INVENTION: TUBERCULOSIS RESISTANCE AGAINST ANTIMICROBIAL REACTIVE
; TITLE OF INVENTION: OXYGEN AND NITROGEN INTERMEDIATES
; NUMBER OF SEQUENCES: 13
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Nixon, Hargrave, Devans & Doyle LLP
; STREET: Clinton Square, P.O. Box 1051
; CITY: Rochester
; STATE: New York
; COUNTRY: USA
; ZIP: 14603
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/067,626
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 60/045,688
; FILING DATE: 06-MAY-1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Goldman, Michael L.
; REGISTRATION NUMBER: 30,727
```

```
; REFERENCE/DOCKET NUMBER: 19603/491
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (716) 263-1304
; TELEFAX: (716) 263-1600
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1244 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cDNA
US-09-067-626-3

Query Match          4.9%; Score 30.2; DB 4; Length 1244;
Best Local Similarity 49.7%; Pred. No. 1.4;
Matches 77; Conservative 0; Mismatches 78; Indels 0; Gaps 0;

QY 185 tgattatgaaaaaatggcctggctgcatacgtgattatgtgtggcggtggcctgtatc 244
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 650 TGTCTAGCGACGAGATCCCGTGGTCCGATCCGCGCTGCGGCGGCTGCGACGCTGCATC 709
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 245 tctgtgtaggtttaccagatgtacatgtgtgcaactgaaaaaaggcggtttctgcac 304
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 710 TCGGGGACCCGTCACGATGGCGCCGCCGAGGAGAGCTGCGCGGAGGCGGACG 769
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 305 ctgcgccacagatcgagctggcgaaaaadtgggcgc 339
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 770 CCGACTGGCGGATGTTGGTGGCGCGCTGTCGCGCAC 804
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 5
US-07-941-651-1
; Sequence 1, Application US/07941651
; Patent No. 532808
; GENERAL INFORMATION:
; APPLICANT: Boston, Rebecca S.
; APPLICANT: Bass, Henry W.
; APPLICANT: O'Brian, Gregory R.
; TITLE OF INVENTION: DNA Encoding A Ribosome Inactivating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 532808th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/941,651
; FILING DATE: 19920908
; CLASSIFICATION: 800
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1934 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
```

```
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 990..1826
US-07-941-651-1

Query Match          4.7%; Score 29; DB 1; Length 1934;
Best Local Similarity 48.0%; Pred. No. 4.6;
Matches 83; Conservative 0; Mismatches 90; Indels 0; Gaps 0;

QY 218 tgattatgtggcggtggcctgtatctctgtctgggatgtttaccagatgtctacgtgtg 277
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1576 TGATCATGTTATGCGAGGGGCTGCGGTCTTTCACCGTGTCCCGCAAGGTAGACGAGGSGT 1635
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 278 cactgaaaaaaggcggtttctgcaacctgcacacagctcgagctgcgaaaaagtgggc 337
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1636 TCAAGAAGCCGCAAGCGGTGACCATATCGCGCTGGAGGGGGAAGCAGGTGCAGAAATGGG 1695
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
QY 338 gcagtttcttgaaggtttactgaccaatctcgtaataatccgaaagcgattatc 390
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1696 ACAGGATCTCGAAAGCGGTCTTCAGGTGGGCCGCTGACCCGACCGCTGAGATC 1748
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 6
US-08-279-996-1
; Sequence 1, Application US/08279996
; Patent No. 5552140
; GENERAL INFORMATION:
; APPLICANT: Boston, Rebecca S.
; APPLICANT: Bass, Henry W.
; APPLICANT: O'Brian, Gregory R.
; TITLE OF INVENTION: DNA Encoding A Ribosome Inactivating
; TITLE OF INVENTION: Protein
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Kenneth D. Sibley; Bell, Seltzer, Park and
; STREET: Post Office Drawer 34009
; CITY: Charlotte
; STATE: No. 5552140th Carolina
; COUNTRY: U.S.A.
; ZIP: 28234
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/279,996
; FILING DATE: 25-JUL-1994
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/941,651
; FILING DATE: 08-SEP-1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Sibley, Kenneth D.
; REGISTRATION NUMBER: 31,665
; REFERENCE/DOCKET NUMBER: 5051-199
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 919-881-3140
; TELEFAX: 919-881-3175
; TELEX: 575102
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1934 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 990..1826
```


Oy 266 tgctacgtggtgcactgaaaaa 287
|| | | | | | | | | | |
Db 5664 AACTGTGTTATTTTCTAAGAA 5685

Search completed: May 6, 2001, 16:25:06
Job time: 2647 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 6, 2001, 14:41:24 ; Search time 1151.46 Seconds
(without alignments)
4688.750 Million cell updates/sec

Title: US-09-466-935-3_COPY_187_804
Perfect score: 618
Sequence: 1 atgttgatgtatttctcac.....ttcatttgattattcggg 618

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 9623517 seqs, 4368049070 residues
Total number of hits satisfying chosen parameters: 19247034

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1:	gb_est1.*	44:	em_esthum10.*
2:	gb_est2.*	45:	em_esthum11.*
3:	gb_est3.*	46:	em_esthum12.*
4:	gb_est4.*	47:	em_esthum13.*
5:	gb_est5.*	48:	em_esthum14.*
6:	gb_est6.*	49:	em_esthum15.*
7:	gb_est7.*	50:	em_esthum16.*
8:	gb_est8.*	51:	em_esthum17.*
9:	gb_est9.*	52:	em_esthum18.*
10:	gb_est10.*	53:	em_esthum19.*
11:	gb_est11.*	54:	em_esthum20.*
12:	gb_est12.*	55:	em_esthum21.*
13:	gb_est13.*	56:	em_esthum22.*
14:	gb_est14.*	57:	em_esthum23.*
15:	gb_est15.*	58:	em_esthum24.*
16:	gb_est16.*	59:	em_esthum25.*
17:	gb_est17.*	60:	em_esthum26.*
18:	gb_est18.*	61:	em_esthum27.*
19:	gb_est19.*	62:	em_esthum28.*
20:	gb_est20.*	63:	em_estin1.*
21:	gb_est21.*	64:	em_estin2.*
22:	gb_est22.*	65:	em_estin3.*
23:	gb_est23.*	66:	em_estin4.*
24:	gb_est24.*	67:	em_estin5.*
25:	gb_est25.*	68:	em_estomi.*
26:	gb_est26.*	69:	em_estom2.*
27:	gb_est27.*	70:	em_estov1.*
28:	gb_est28.*	71:	em_estov2.*
29:	gb_est29.*	72:	em_estpl1.*
30:	gb_est30.*	73:	em_estpl2.*
31:	gb_est31.*	74:	em_estpl3.*
32:	gb_est32.*	75:	em_estpl4.*
33:	gb_est33.*	76:	em_estpl5.*
34:	gb_est34.*	77:	em_estpl6.*
35:	gb_est35.*	78:	em_estpl7.*
36:	gb_est36.*	79:	em_estpl8.*
37:	gb_est37.*	80:	em_estpl9.*
38:	gb_est38.*	81:	em_estpl10.*
39:	gb_est39.*	82:	em_estro1.*
40:	gb_est40.*	83:	em_estro2.*
41:	em_estba.*	84:	em_estro3.*
42:	em_estfun.*	85:	em_estro4.*
43:	em_esthum1.*	86:	em_estro5.*
44:	em_esthum2.*	87:	em_estro6.*
45:	em_esthum3.*	88:	em_estro7.*
46:	em_esthum4.*	89:	em_estro8.*
47:	em_esthum5.*	90:	em_estro9.*
48:	em_esthum6.*	91:	em_estro10.*
49:	em_esthum7.*	92:	em_estro11.*
50:	em_esthum8.*	93:	em_estro12.*
51:	em_esthum9.*	94:	em_estro13.*
52:		95:	em_estro14.*
53:		96:	em_estro15.*
54:		97:	em_estro16.*
55:		98:	em_estro17.*
56:		99:	em_estro18.*
57:		100:	em_estro19.*
58:		101:	em_estro20.*
59:		102:	gb_est25.*
60:		103:	gb_est26.*
61:		104:	gb_est27.*
62:		105:	gb_est28.*
63:		106:	gb_est29.*
64:		107:	gb_est30.*
65:		108:	gb_est31.*
66:		109:	gb_est32.*
67:		110:	gb_est33.*
68:		111:	gb_est34.*
69:		112:	gb_est35.*
70:		113:	gb_est36.*
71:		114:	gb_est37.*
72:		115:	gb_est38.*
73:		116:	gb_est39.*

117: gb_est48:*
118: gb_est49:*
119: gb_est50:*
120: gb_est51:*
121: gb_est52:*
122: gb_est53:*
123: gb_est54:*
124: gb_est55:*
125: gb_est56:*
126: gb_est57:*
127: gb_est58:*
128: gb_est59:*
129: gb_est60:*
130: gb_est61:*
131: gb_est62:*
132: gb_est63:*
133: gb_est64:*
134: gb_est65:*
135: gb_est66:*
136: gb_est67:*
137: gb_est68:*
138: gb_est69:*
139: gb_est70:*
140: gb_est71:*
141: gb_est72:*
142: gb_est73:*
143: gb_est74:*
144: gb_est75:*
145: gb_est76:*
146: gb_est77:*
147: gb_est78:*
148: gb_est79:*
149: gb_est80:*
150: gb_est81:*
151: gb_est82:*
152: gb_est83:*
153: gb_est84:*
154: gb_est85:*
155: gb_est86:*
156: gb_est87:*
157: gb_est88:*
158: gb_est89:*
159: gb_est90:*
160: gb_est91:*
161: gb_est92:*
162: gb_est93:*
163: gb_est94:*
164: gb_est95:*
165: gb_est96:*
166: gb_est97:*
167: gb_est98:*
168: gb_est99:*
169: gb_est100:*
170: gb_est101:*
171: gb_est102:*
172: gb_est103:*
173: gb_est104:*
174: gb_est105:*
175: gb_est106:*
176: gb_est107:*
177: gb_est108:*
178: gb_est109:*
179: gb_est110:*
180: gb_est111:*
181: gb_est112:*
182: gb_est113:*
183: gb_est114:*
184: gb_est115:*
185: gb_est116:*
186: gb_est117:*
187: gb_est118:*
188: gb_est119:*
189: gb_est120:*

190: em_gss_pln1:*
191: em_gss_pln2:*
192: em_gss_pro:*
193: em_gss_rod1:*
194: em_gss_rod2:*
195: em_gss_rod3:*
196: em_gss_rod4:*
197: em_gss_rod5:*
198: em_gss_vrt1:*
199: em_gss_vrt2:*
200: em_gss_vrt3:*
201: gb_gss1:*
202: gb_gss2:*
203: gb_gss3:*
204: gb_gss4:*
205: gb_gss5:*
206: gb_gss6:*
207: gb_gss7:*
208: gb_gss8:*
209: gb_gss9:*
210: gb_gss10:*
211: gb_gss11:*
212: gb_gss12:*
213: gb_gss13:*
214: gb_gss14:*
215: gb_gss15:*
216: gb_gss16:*
217: gb_gss17:*
218: gb_gss18:*
219: gb_gss19:*
220: gb_gss20:*
221: gb_gss21:*
222: gb_gss22:*
223: gb_gss23:*
224: gb_gss24:*
225: gb_gss25:*
226: gb_gss26:*
227: gb_gss27:*
228: gb_gss28:*
229: gb_gss29:*
230: gb_gss30:*
231: gb_gss31:*
232: gb_gss32:*
233: gb_gss33:*
234: gb_gss34:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match		Query Length	DB ID	Description
		No.	Match			
1	38.6	6.2	922	229	CNS0073W	AL066784 Drosophil
2	38.4	6.2	425	204	AQ280480	AQ280480 C11B1-E1-
3	36.2	5.9	1108	139	BE793235	BE793235 601583674
4	35.8	5.8	977	229	CNS00JX7	AL076850 Drosophil
5	35.6	5.8	912	229	CNS006RB	AL065826 Drosophil
6	35.6	5.8	939	229	CNS00CNG	AL059400 Drosophil
7	35.2	5.7	914	231	CNS0300A	AL252739 Tetraodon
8	35.2	5.7	1016	231	CNS04744	AL277501 Tetraodon
9	35.2	5.7	1068	230	CNS028DO	AL185829 Tetraodon
10	35	5.7	1200	229	CNS01671	AL106392 Drosophil
11	34.6	5.6	835	150	BF621090	BF621090 HVSMea000
12	34.6	5.6	920	229	CNS0062R	AL061710 Drosophil
13	34.4	5.6	1101	229	CNS017GJ	AL108013 Drosophil
14	34.2	5.5	832	142	BE957648	BE957648 601653724
15	33.8	5.5	501	32	AV720117	AV720117 AV720117
16	33.6	5.4	268	119	AW721077	AW721077 833011D11
17	33.6	5.4	1355	164	BE196650	BE196650 HVSMea009
18	33.4	5.4	571	206	AQ438206	AQ438206 HS_5083_B

C	19	33.4	5.4	615	118	AW678708
C	20	33.4	5.4	692	118	AW678638
C	21	33.4	5.4	703	118	AW678815
C	22	33.4	5.4	737	205	AQ350746
C	23	33.4	5.4	895	229	CNS0071A
C	24	33.2	5.4	488	19	A1384522
C	25	33.2	5.4	579	24	A1730513
C	26	33.2	5.4	837	165	BE284012
C	27	33.2	5.4	1101	229	CNS00KK2
C	28	33.2	5.4	1772	119	AW729450
C	29	33	5.3	283	146	BF273544
C	30	32.8	5.3	284	124	BB089204
C	31	32.8	5.3	525	205	AQ366581
C	32	32.8	5.3	884	229	CNS006U0
C	33	32.8	5.3	993	231	CNS04DR2
C	34	32.6	5.3	739	172	BG030901
C	35	32.6	5.3	743	138	BF729153
C	36	32.6	5.3	869	232	CNS051EH
C	37	32.6	5.3	1042	172	CNS05D89
C	38	32.6	5.2	470	17	A1168689
C	39	32.4	5.2	499	24	A1781974
C	40	32.4	5.2	672	167	BE459806
C	41	32.4	5.2	859	144	BF107761
C	42	32.4	5.2	1101	229	CNS00L10
C	43	32.2	5.2	310	228	B98159
C	44	32.2	5.2	481	209	AQ627446
C	45	32.2	5.2	483	32	AV179757

ALIGNMENTS

RESULT	1
CNS0073W	922 bp DNA GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TET3 end of BAC #
DEFINITION	BACR14D09 of RPCI-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.
ACCESSION	AL066784
VERSION	AL066784.1 GI:4945247
KEYWORDS	GSS.
SOURCE	fruit fly.
ORGANISM	Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
AUTHORS	1 (bases 1 to 922)
TITLE	Genoscope.
JOURNAL	Direct Submission
COMMENT	Submitted (02-JUN-1999) Genoscope - Centre National de Sequences : BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr - Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACs. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mammoser in Pictet de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BDGP's PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm . Location/Qualifiers 1. 922 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACR14D09"
FEATURES	source

BASE COUNT	223 a	95 c	109 g	221 t	274 others
ORIGIN	/note="end : TET3"				
Query Match	6.2%; Score 38.6; DB 229; Length 922;				
Best Local Similarity	20.7%; Pred.No.0.37; Indels 0; Gaps 0;				
Matches	47; Conservative 83; Mismatches 97;				
Qy	387	tatactatttggctcgtgttctcattgtttgtcgtgtatacaactgttgccactacgcgcg	446		
Db	610	TGKKTGCGKTKTKTKTKKGGTKKKKTKTKGKKGCTGKGGTKTGTGGKKTKTKTKKKK	669		
Qy	447	ctggggcattttggcctgaatcgtgcgaacgcgtgcgctgggtttaccgcctgttgcag	506		
Db	670	KTGKKKKKTKGKGGKTKTKTKTKGKGGKTKKKTKTKTKTKTKKKKKTKTKT	729		
Qy	507	ccctgttgcctcgcgcgaactgcgcgtgtgttatacaactgtctgcgaagtgcattgatgg	566		
Db	730	KKTGTGKKKTKTKTKTKGKGTGKGGKGGKTKGTGTKGKTGGKGGKGGKGGKGGKGGK	789		
Qy	567	ttttgcggcggttattttgcgcgattttggcattcattgattatt	613		
Db	790	GKKKTKGKGTGKTKTKGKTKTKTKKKKKKKTKTKKKGTKGSGT	836		
RESULT	2				
AQ280480/c	AQ280480 425 bp DNA GSS 22-NOV-1998				
LOCUS	CITBI-EI-2516B16.TF CITBI-EI Homo sapiens genomic clone 2516B16,				
DEFINITION	DNA sequence.				
ACCESSION	AQ280480				
VERSION	AQ280480.1 GI:3906299				
KEYWORDS	GSS.				
SOURCE	human.				
ORGANISM	Homo sapiens				
REFERENCE	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.				
AUTHORS	Adams,M.D., Rounsley,S.D., Zhao,S., Bass,S., Linher,K., Golden,K., Berry,K., Grainger,D., Suh,E., Wible,C., Shizuya,H., Simon,M. and Venter,J.C.				
TITLE	Use of a random human BAC End Sequence Database for Sequence-Ready Map Building				
JOURNAL	Unpublished (1998)				
COMMENT	Other_GSSs: CITBI-EI-2516B16.TR Contact: Mark Adams Department of Eukaryotic Genomics The Institute for Genomic Research 9712 Medical Center Dr., Rockville, MD 20850, USA Tel: 301 838 0200 Fax: 301 838 0208 Email: mdadams@tigr.org Clones are available from Research Genetics (info@resgen.com). BAC end search page: http://www.tigr.org/tldb/humgen/bac_end_search/bac_end_search.html. Seq primer: M13-21 Class: BAC ends.				
FEATURES	Location/Qualifiers				
source	1. .425 /organism="Homo sapiens" /db_xref="taxon:9606" /clone="2516B16" /clone_lib="CITBI-EI" /sex="male" /cell_type="sperm" /note="Vector: pBeloBAC11; Site_1: EcoRI; Site_2: EcoRI; Caltech Human BAC Library D"				
BASE COUNT	111 a	140 c	100 g	74 t	
ORIGIN					

Query Match

```
Best Local Similarity 49.08; Pred. No. 0.35;
Matches 102; Conservative 0; Mismatches 106; Indels 0; Gaps 0;

QY 119 tggcgctgctggcattaccctgcgcgtaattggttggtggctgggattgcgctgtggcc 178
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 241 TGGGCTAGCCAGGATCGCTGGTGGGAGGATAGCTGCACTGGGGGGGATGTGAGGGTT 182
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 179 tcatattgattcgaataaaatggcctgcctgcatacgcctgattatgtggcggtggcc 238
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 181 GCATTGCGAGACCTTCCATCGCGCTGCCATCTGCGGCCAGAGATTGGCCCTGGGC 122
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 239 tgtatctctgtggtggtggtaccagatgctacgtgtgctacgtgaaagagcggttt 298
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 121 TCTGCTGTGGTCTCGTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTGTGCTG 62
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 299 ctgcacctgcgcacagctcagctgccc 326
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 61 CTCACCTTACACTCCACCCCTTCAGGC 34
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 3
LOCUS BE793235 1108 bp mRNA EST 20-SEP-2000
DEFINITION 601583674F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3937891 5',
mRNA sequence.
ACCESSION BE793235
VERSION BE793235.1 GI:10214433
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1108)
AUTHORS NIH-MGC http://mgi.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert.Strausberg@nih.gov
Tissue Procurement: DCTD/DTF
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: image.llnl.gov
plate: L1CM784 row: d column: 20
High quality sequence stop: 719.
Location/Qualifiers
FEATURES
source
1..1108
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:3937891"
/clone_lib="NIH_MGC_7"
/tissue_type="small cell carcinoma"
/cell_line="WGC3"
/lab_host="DH10B (phage-resistant)"
/note="Organ: lung; Vector: pOTB7; Site_1: XhoI; Site_2:
EcoRI; cDNA made by oligo-dT priming. Directionally
cloned into EcoRI/XhoI sites using the following 5',
adaptor: GGCACGAG(G). Size-selected >500bp for average
insert size 1.8kb. Library constructed by Ling Hong in
the laboratory of Gerald M. Rubin (University of
California, Berkeley) using ZAP-cDNA synthesis kit
(Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 334 a 284 c 310 g 180 t

```
Query Match 5.98; Score 36.2; DB 139; Length 1108;
Best Local Similarity 54.08; Pred. No. 2.2;
Matches 74; Conservative 0; Mismatches 63; Indels 0; Gaps 0;

QY 48 tatgagcccggtcccgattcttttttctctcagacgcgtgctgcagtgtaacgtaa 107
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
```

```
Db 979 TATGTGTCGGCTCAGCCCTCTGGGTTCGCTCCAGCCCTCTCGGTGTCGTCTCGTGT 920
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 108 agaagcgtatgtagggcgctgctgggcattacctgcgcgtaattggttggtggctgggattgc 167
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 919 TTTGGGATCTCTTGATATGTTGGCTTTATCGCGCGCTTATATCTTCTTCTTCTGACCTCC 860
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 168 gctgctggcctgcatt 184
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 859 GTCTCTAGCCCGCGCTT 843
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 4
LOCUS CNS00JX7 977 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence TET3 end of BAC:
BACR39C01 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.
ACCESSION AL076850
VERSION AL076850.1 GI:4956428
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
REFERENCE 1 (bases 1 to 977)
AUTHORS Genoscope.
TITLE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT Determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information
please see http://www.fruitfly.org The BDGP Drosophila
melanogaster BAC library was prepared by Kazutoyo Osoegawa and
Aaron Mamoser in Pieter de Jong's laboratory in the Department of
Cancer Genetics at the Roswell Park Cancer Institute in Buffalo,
NY. The library is named RPCI-98 and was constructed by partial
EcoRI digestion of Drosophila DNA provided by the BDGP from the
isogenic strain Y2; cn bw sp, the same strain used for the BDGP's
P1 and EST libraries. A more detailed description of the library
and how to order individual BAC clones, the entire library, or
filters for hybridization from the BACPAC Resource Center can be
found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
```

FEATURES

source
1..977
/organism="Drosophila melanogaster"
/db_xref="taxon:7227"
/clone_lib="RPCI-98"
/clone="BACR39C01"
/note="end : TET3"

BASE COUNT 132 a 80 c 110 g 263 t 392 others

ORIGIN

Query Match 5.88; Score 35.8; DB 229; Length 977;
Best Local Similarity 17.18; Pred. No. 2.9;
Matches 22; Conservative 66; Mismatches 41; Indels 0; Gaps 0;

QY 112 gcgatgagggcgctgctgggcattacctgcgcgtaattggttggtggctgggattgcgctg 171
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 154 SSGATSSSSSSGSSGSSSTATTSSSTSSSTSSSTSSSTSSSTSSSTSSSTSSSB 213
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 172 ctggcctgcattgattatcgaaaaatggcctggctgcatacgcgtgattggtgggc 231
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 214 SSSSSSSSSSTSSSTTSVSSSATTTBSSBSSSSASBSSSSSSSTSSSTSSS 273
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

QY 232 ggtggcctg 240
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Db 274 SSTBSSTS 282
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

Holacanthopterygii; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE
AUTHORS 1 (bases 1 to 914)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizes,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 914)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL
REFERENCE Unpublished
AUTHORS 3 (bases 1 to 914)
Genoscope.
Direct Submission
TITLE
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/tetraodon.

FEATURES
source
Location/Qualifiers
1..914
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="04IG19"
/clone.lib="G"
/note="Genoscope sequence ID : COBG041AD10LP1-end : T7"
BASE COUNT 201 a 242 C 293 g 175 t 3 others
ORIGIN

Query Match 5.7%; Score 35.2; DB 231; Length 914;
Best Local Similarity 54.7%; Pred. No. 4.3;
Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 433 ggcactaccgcgcgtgggcatattttgcgtgatcattgtcgaaacgctggcggttt 492
Db 874 GGTTCACCGCGCGCACACGCATTCTCGCTCAGACGCCCATAAACCTCGCGCTCAG 815

QY 493 accgtgttgcagcgtgtttgcctgcgcgaatgcgcgtgttatcaacgtctggcg 552
Db 814 AGGGGTGTCGCGTACCCAGCGCTGAGCGCGCTTCACAGCGCGCGCTAG 755

QY 553 aagtgagat 560
Db 754 ACGTGCGT 747

RESULT 8
CNS04744
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence T7 end of clone
086H24 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL277501
VERSION
AL277501.1 GI:8011704
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
Holacanthopterygii; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodon.
1 (bases 1 to 1016)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizes,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE
Characterization and repeat analysis of the compact genome of the

freshwater pufferfish Tetraodon nigroviridis

JOURNAL
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 1016)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Human gene number estimate provided by genome wide analysis using
Tetraodon nigroviridis DNA sequence

JOURNAL
REFERENCE Unpublished
AUTHORS 3 (bases 1 to 1016)
Genoscope.
Direct Submission
TITLE
Submitted (12-APR-2000) to the EMBL/GenBank/DBJ databases
JOURNAL
COMMENT This sequence is a single read and was generated as part of a large
scale clone-end sequencing project of the Tetraodon nigroviridis
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetraodon.

FEATURES
source
Location/Qualifiers
1..1016
/organism="Tetraodon nigroviridis"
/db_xref="taxon:99883"
/clone="086H24"
/clone.lib="G"
/note="Genoscope sequence ID : COBG086DD12LP1-end : T7"
BASE COUNT 238 a 295 C 269 g 212 t 2 others
ORIGIN

Query Match 5.7%; Score 35.2; DB 231; Length 1016;
Best Local Similarity 54.7%; Pred. No. 4.5;
Matches 70; Conservative 0; Mismatches 58; Indels 0; Gaps 0;

QY 433 ggcactaccgcgcgtgggcatattttgcgtgatcattgtcgaaacgctggcggttt 492
Db 573 GGCTTCACCGCGCGCACACGCATTCTCGCTCAGACGCCCATAAACCTCGCGCTCAG 632

QY 493 accgtgttgcagcgtgtttgcctgcgcgaatgcgcgtgttatcaacgtctggcg 552
Db 633 AGGGGTGTCGCGTACCCAGCGCTGAGCGCGCTTCACAGCGCGCGCTAG 692

QY 553 aagtgagat 560
Db 693 ACGTGCGT 700

RESULT 9
CNS028DO
LOCUS
DEFINITION
Tetraodon nigroviridis genome survey sequence PUC-Ori end of clone
244L11 of library G from Tetraodon nigroviridis, genomic survey
sequence.
ACCESSION
AL185829
VERSION
AL185829.1 GI:7823933
KEYWORDS
GSS; genome survey sequence.
SOURCE
Tetraodon nigroviridis.
ORGANISM
Tetraodon nigroviridis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Euacanthomorpha;
Holacanthopterygii; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodon.
1 (bases 1 to 1068)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Fizes,C., Fisher,C.,
Bouneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.
TITLE
Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetraodon nigroviridis

JOURNAL
REFERENCE Unpublished
AUTHORS 2 (bases 1 to 1068)
Roest-Crollius,H., Jaillon,O., Dasilva,C., Bouneau,L., Fisher,C.,
Bernot,A., Fizes,C., Wincker,P., Brottier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.
TITLE
Human gene number estimate provided by genome wide analysis using

Tetraodon nigroviridis DNA sequence		/clone="BACN15022"		164 a		225 c		251 g		374 t		186 others	
JOURNAL	Unpublished												
REFERENCE	3 (bases 1 to 1068)												
AUTHORS	Genoscope.												
TITLE	Direct Submission												
JOURNAL	Submitted (12-APR-2000) to the EMBL/GenBank/DBDJ databases												
COMMENT	This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis genome. For more information, please take a look at http://www.genoscope.cns.fr/tetraodon.												
FEATURES		Location/Qualifiers											
source	1..1068	/organism="Tetraodon nigroviridis"											
		/db_xref="taxon:99883"											
		/clone="244L11"											
		/clone_lib="G"											
		/note="Genoscope sequence ID : C0AG244CF06SPl-end : PUC-Ori"											
BASE COUNT	251 a	301 c	272 g	237 t	7 others								
ORIGIN													
Query Match		5.7%; Score 35.2; DB 230; Length 1068;											
Best Local Similarity		54.7%; Pred.No.4.5;											
Matches		70; Conservative 0; Mismatches 59; Indels 0; Gaps 0;											
QY	433	ggcactacgcgcgtggggcatttttggcgtgattctgtcgaacgcgtggcggttt	492										
Db	829	GGCTTACCGCGGCACAGCATTCCTGCGCTCAGCAGCCCATTAACCTCGCGGTGCAG	888										
QY	493	accgtgttcagcagctgtttccctgcgcgaatgcgcctgtgttatacaacgtctggcg	552										
Db	889	AGGGGTGTCGCGTGCCTTACCCAGCGGTGACGCGCGCTTCACAGCGCGCGGTAG	948										
QY	553	aagtggat 560											
Db	949	ACGTCGCT 956											
RESULT 10													
CNS01671													
LOCUS	CNS01671	1200 bp	DNA	GSS	26-JUL-1999								
DEFINITION	Drosophila melanogaster genome survey sequence SP6 end of BAC BACN15022 of DrosBAC library from Drosophila melanogaster (fruit fly), genomic survey sequence.												
ACCESSION	AL106392												
VERSION	AL106392.1	GI:5621686											
KEYWORDS	GSS.												
SOURCE	fruit fly.												
ORGANISM	Plasmid Drosophila melanogaster												
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.													
REFERENCE	1 (bases 1 to 1200)												
AUTHORS	Genoscope.												
TITLE	Direct Submission												
JOURNAL	Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : seque@genoscope.cns.fr												
COMMENT	- Web : www.genoscope.cns.fr) Determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome Project (EDGP) - http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billaud at CEPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MPC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelOBAC11.												
FEATURES		Location/Qualifiers											
source	1..1200	/organism="Drosophila melanogaster"											
		/plasmid="pBelOBAC11"											
		/db_xref="taxon:7227"											
		/clone_lib="DrosBAC"											

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:38:48 ; Search time 34.23 Seconds
(without alignments)
344.015 Million cell updates/sec

Title: US-09-466-935-4
Perfect score: 1054
Sequence: 1 MLMFLITVAMVHVALMSPG.....IDGFAGALFAGFIHLIISR 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 390729 seqs, 57163235 residues

Total number of hits satisfying chosen parameters: 390729

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : A_Geneseq_0401.*

- 1: /SID52/gcgdata/geneseq/geneseq/AA1980.DAT.*
- 2: /SID52/gcgdata/geneseq/geneseq/AA1981.DAT.*
- 3: /SID52/gcgdata/geneseq/geneseq/AA1982.DAT.*
- 4: /SID52/gcgdata/geneseq/geneseq/AA1983.DAT.*
- 5: /SID52/gcgdata/geneseq/geneseq/AA1984.DAT.*
- 6: /SID52/gcgdata/geneseq/geneseq/AA1985.DAT.*
- 7: /SID52/gcgdata/geneseq/geneseq/AA1986.DAT.*
- 8: /SID52/gcgdata/geneseq/geneseq/AA1987.DAT.*
- 9: /SID52/gcgdata/geneseq/geneseq/AA1988.DAT.*
- 10: /SID52/gcgdata/geneseq/geneseq/AA1989.DAT.*
- 11: /SID52/gcgdata/geneseq/geneseq/AA1990.DAT.*
- 12: /SID52/gcgdata/geneseq/geneseq/AA1991.DAT.*
- 13: /SID52/gcgdata/geneseq/geneseq/AA1992.DAT.*
- 14: /SID52/gcgdata/geneseq/geneseq/AA1993.DAT.*
- 15: /SID52/gcgdata/geneseq/geneseq/AA1994.DAT.*
- 16: /SID52/gcgdata/geneseq/geneseq/AA1995.DAT.*
- 17: /SID52/gcgdata/geneseq/geneseq/AA1996.DAT.*
- 18: /SID52/gcgdata/geneseq/geneseq/AA1997.DAT.*
- 19: /SID52/gcgdata/geneseq/geneseq/AA1998.DAT.*
- 20: /SID52/gcgdata/geneseq/geneseq/AA1999.DAT.*
- 21: /SID52/gcgdata/geneseq/geneseq/AA2000.DAT.*
- 22: /SID52/gcgdata/geneseq/geneseq/AA2001.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1054	100.0	206	21	E. coli L-threonin
2	261.5	24.8	223	21	Escherichia coli Y
3	190	18.0	212	21	Escherichia coli Y
4	179.5	17.0	229	21	A. vitis hyperiens
5	122	11.6	205	21	E. coli L-homoseri
6	122	11.6	205	21	E. coli RhtB prote
7	105	10.0	236	18	C. glutamicum Lys
8	103	9.8	211	21	Escherichia coli Y
9	101	9.6	195	21	Escherichia coli Y
10	94.5	9.0	130	21	Human ORFX ORF171
11	91	8.6	153	17	Mel-linked mlga ge

12	91	8.6	153	20	W73358	S. colwelliana Mlg
13	86.5	8.2	431	17	R98004	Oestrogen-regulate
14	86.5	8.2	431	19	R34528	Protein encoded by
15	86	8.2	390	11	R05475	Chloramphenicol re
16	82.5	7.8	356	20	Y38893	Neisseria gonorrhoe
17	82	7.8	713	21	Y58582	Sorangium cellulos
18	81.5	7.7	439	21	Y39079	Arabidopsis thalia
19	81.5	7.7	444	21	G21141	Arabidopsis thalia
20	81.5	7.7	444	21	G39078	Arabidopsis thalia
21	81.5	7.7	555	21	G39077	Arabidopsis thalia
22	81.5	7.7	559	21	G21140	Arabidopsis thalia
23	81.5	7.7	579	21	G21139	Arabidopsis thalia
24	81.5	7.7	1170	20	W88447	Yeast NPC1 protein
25	81	7.7	381	11	R03463	GPL transglutamina
26	80.5	7.6	299	20	W90237	E. coli antibiotic
27	80.5	7.6	306	20	W90236	E. coli antibiotic
28	80	7.6	356	20	Y38890	Neisseria meningit
29	80	7.6	356	20	Y38891	Neisseria meningit
30	80	7.6	609	21	B08549	Amino acid sequenc
31	79.5	7.5	1307	17	R99255	Aspergillus flavus
32	78	7.4	1518	20	Y16099	Acetobacter xylinu
33	77	7.3	332	20	Y34112	Human GABA recepto
34	77	7.3	332	20	Y30312	Human GABA recepto
35	77	7.3	332	20	Y29662	Amino acid sequenc
36	77	7.3	859	20	Y34111	Human 7TM receptor
37	77	7.3	859	20	Y30311	Human GABA recepto
38	77	7.3	874	20	Y29661	Amino acid sequenc
39	77	7.3	898	20	Y14082	Human 7TM receptor
40	77	7.3	898	20	Y70326	Human GABABR2 prot
41	77	7.3	914	21	Y44344	Human gamma amino
42	77	7.3	941	20	Y29796	Protein-1 related
43	77	7.3	941	21	W90938	Human gamma-amino-
44	77	7.3	941	21	Y51928	Human GABA-B recep
45	77	7.3	941	21	Y70328	Human GABA-B-R2 re

ALIGNMENTS

RESULT	1
Y99598	ID Y99598 standard; Protein; 206 AA.
XX	XX
AC	Y99598;
XX	XX
XX	XX
DE	08-SEP-2000 (first entry)
XX	XX
DE	E. coli L-threonine resistance protein, RhtC.
XX	XX
XX	L-threonine resistance; L-threonine synthesis; rhtC;
KW	L-homoserine; L-valine; L-leucine.
XX	XX
OS	Escherichia coli.
XX	XX
PN	EP1013765-A1.
XX	XX
PD	28-JUN-2000.
XX	XX
PF	20-DEC-1999; 99EP-0125406.
XX	XX
PR	23-DEC-1998; 98RU-0123511.
XX	XX
PA	(AJIN) AJINOMOTO KK.
XX	XX
PI	Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;
XX	XX
DR	WPI; 2000-414602/36.
DR	N-PSDB; A48443.
XX	XX
PT	Novel Escherichia bacterium having enhanced L-threonine resistance due
PT	to enhanced RhtC protein activity, used to produce L-threonine,
PT	L-homoserine, L-valine and L-leucine -
XX	XX

PS Claim 1; Page 15-16; 24pp; English.

XX The present sequence is the L-threonine resistance protein, RhtB, from
 CC Escherichia coli. The coding sequence may be used to impart L-threonine
 CC resistance on E. coli bacteria, which would be useful for producing
 CC a high yield of L-threonine. L-threonine resistance means that the
 CC bacteria will be able to grow on a minimal medium containing
 CC L-threonine at a concentration at which the corresponding wild-type
 CC strain would not grow. Since the transformed bacteria can grow on the
 CC minimal medium, it can synthesize L-threonine, which accumulates. The
 CC accumulated amino acids can then be removed from the culture medium.
 CC The bacterium of the present invention may also be used to synthesize
 CC L-homoserine, L-valine and L-leucine at increased levels.

SQ Sequence 206 AA;

Query Match 100.0%; Score 1054; DB 21; Length 206;
 Best Local Similarity 100.0%; Pred. No. 1.9e-112;
 Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MMLFLTVMVHIVAMSPGDPFFVSTAVSRKKEAMGVLTGCGVMVWAGIALGL 60
 DB 1 mmlfltvmvhiavmspgdpffvstavsrsrkeamgvlgtcgvmvwagialgl 60
 QY 61 HLIIEKMAWLHTLIMVGGGLYLCWMGYOMLRGALKKEAVSAPQVELAKSGRFLKGL 120
 DB 61 hliiekmawlhtlimvgggllylcwmgyomlrgalkkeavsapqvelaksgrsflkgl 120
 QY 121 TNLAMPKAIIFGVSFLFVDGNTGTTARWGFALIIIVETLAWFTVVASLFPQMRRGY 180
 DB 121 tnlampkaiifgvsflfvdgntgttarwgfalfalivetlawftvvaslfpqmrgrgy 180
 QY 181 QRLAKWIDGFAGALPAGFGIHLITSR 206
 DB 181 qrlakwidgfagalfagfgihliisr 206

RESULT 2
 B01786
 ID B01786 standard; Protein; 223 AA.
 XX
 AC B01786;

DT 03-JAN-2001 (first entry)
 DE Escherichia coli Yahn amino acid excretion protein.
 XX
 KW E. coli; yahn gene; amino acid production; excretion protein gene;
 KW amino acid excretion protein.

OS Escherichia coli.

PN EP1016710-A2.

PD 05-JUL-2000.

PF 17-DEC-1999; 99EP-0125263.

PR 30-DEC-1998; 98RU-0124016.

PR 09-MAR-1999; 99RU-0104431.

XX (AJIN) AJINOMOTO CO INC.

XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
 PI Tokhmakova IL;

DR WPI; 2000-414802/36.

XX N-PSDB; A52688.

PT Increased production of L-amino acids by an Escherichia bacterium
 PT comprises increasing the expression amount of an L-amino acid excretion
 PT protein -

XX Claim 1; Page 19; 29pp; English.

XX The present sequence is the Yahn amino acid excretion protein from
 CC Escherichia coli. This protein is involved in the production of
 CC amino acids, and an increase in its expression leads to an increase in
 CC accumulation of amino acids in the cell. In this case, an increase in
 CC lysine, glutamic acid and proline is achieved if multiple copies of its
 CC gene are transfected into a bacterium. The bacterium used is E. coli.

SQ Sequence 223 AA;

Query Match 24.8%; Score 261.5; DB 21; Length 223;
 Best Local Similarity 28.2%; Pred. No. 6.5e-22;
 Matches 58; Conservative 50; Mismatches 83; Indels 15; Gaps 3;

QY 4 LFTVAMVHIVAMSPGDPFFVSTAVSRKKEAMGVLTGCGVMVWAGIALGLHLI 63
 DB 20 vyltavl-fvitfnpganlfvvvqtsiasgragvltglgvaldafysglglfglati 78
 QY 64 IEKMAWLHTLIMVGGGLYLCWMGYOMLRGALKKEAVSAPQVELAKSGRS-----FLK 117
 DB 79 itceeiifslirivgagylwfwesmr-----rqstpqmstlqqpisapwyvffir 130
 QY 118 GLLTNLAMPKAIIFGVSFLFVDGNTGTTARWGFALIIIVETLAWFTVVASLFPQMR 177
 DB 131 glitdlnspqtlvffisfvtnaetptwarlmawagivlasliiwrvfslsqafslpavr 190
 QY 178 RGYORLAKWIDGFAGALPAGFGIHLI 203
 DB 191 raygrmqvrasvrigaigvafalrli 216

RESULT 3
 B01787
 ID B01787 standard; Protein; 212 AA.
 XX
 AC B01787;

DT 03-JAN-2001 (first entry)

DE Escherichia coli Yeas amino acid excretion protein.

XX E. coli; yeas gene; amino acid production; excretion protein gene;
 KW amino acid excretion protein.

OS Escherichia coli.

PN EP1016710-A2.

PD 05-JUL-2000.

PF 17-DEC-1999; 99EP-0125263.

PR 30-DEC-1998; 98RU-0124016.

PR 09-MAR-1999; 99RU-0104431.

XX (AJIN) AJINOMOTO CO INC.

XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
 PI Tokhmakova IL;

DR WPI; 2000-414802/36.

XX N-PSDB; A52689.

PT Increased production of L-amino acids by an Escherichia bacterium
 PT comprises increasing the expression amount of an L-amino acid excretion
 PT protein -

XX Claim 1; Page 21; 29pp; English.

XX The present sequence is the Yeas amino acid excretion protein from

XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
PI Tokhmakova IL;
XX WPI; 2000-414802/36.
XX N-PSDB; A52690.
DR Increased production of L-amino acids by an *Escherichia* bacterium
PT comprises increasing the expression amount of an L-amino acid excretion
PT protein -
XX Claim 1; Page 23; 29pp; English.
XX The present sequence is the yfik amino acid excretion protein from
CC *Escherichia coli*. This protein is involved in the production of amino
CC acids, and an increase in its expression leads to an increased
CC accumulation of amino acids in the cell. In this case, an increase in
CC threonine, histidine, glutamic acid and proline is achieved if multiple
CC copies of its gene are transfected into a bacterium. The bacterium used
CC is *E. coli*.
XX Sequence 195 AA;
SQ
Query Match 9.6%; Score 101; DB 21; Length 195;
Best Local Similarity 22.0%; Pred. No. 0.0012;
Matches 39; Conservative 32; Mismatches 72; Indels 34; Gaps 7;
QY 1 MMLFLFTVAWHI VALMSPGDPFFVSOTAVSRKEMGVLTGCG----VMWAGIA 56
DB 5 llsafwtv---liamtppnnllalsatsatshgrstgrvlagmslglvilmilcagis 61
QY 57 LGLHLHIEKMAWHLHLMVGGGLYLCVMGYQMLRGALKKEAVSAPAPQVELAKSGRSFL 116
DB 62 ----flavidpaavh-llswagaayivlwakiatsptkedglq-----kpsifw 108
QY 117 KGLTLNLANPKAIIYFGSVFLFVGDNDVGTARWGIFALIIIVETLAWFTVVASLFAL 173
DB 109 asfalqfvnkvilly-----gvta-1stfvlpqatqalswvgvsvllam 151
RESULT 10
ID B40407 standard; Protein; 130 AA.
XX B40407;
XX 08-FEB-2001 (first entry)
DE Human ORFX ORF171 polypeptide sequence SEQ ID NO:342.
XX Human; open reading frame; ORFX; detection; cytostatic; hepatotropic;
KW vulnary; antiparasitic; antiparkinsonian; nootropic; neuroprotective;
KW immunostimulant; osteopathic; antiarthritic; vasotrophic; antidiabetic;
KW hypotensive; dermatological; immunosuppressive; antiinflammatory;
KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
KW cholesterol ester storage; systemic lupus erythematosus; infection;
KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
KW thrombosis; contraceptive.
XX Homo sapiens.
XX WO200058473-A2.
XX 05-OCT-2000.
XX 31-MAR-2000; 2000WO-US08621.

XX 31-MAR-1999; 99US-0127607.
PR 02-APR-1999; 99US-0127636.
PR 05-APR-1999; 99US-0127728.
PR 30-MAR-2000; 2000US-0540763.
XX (CURA-) CURAGEN CORP.
XX Shimkets RA, Leach M;
XX WPI; 2000-602362/57.
DR N-PSDB; C74616.
XX Novel nucleic acids and peptides derived from open reading frame X,
PT useful for treating e.g. cancers, proliferative disorders,
PT neurodegenerative disorders and cardiovascular disease -
XX Claim 11; Page 608; 5507pp; English.
CC C74446 to C77606 encode the proteins given in B40237 to B43397, which
CC represent the human ORFX open reading frames 1 to 3161. The ORFX
CC sequences have activities such as: cytostatic; hepatotropic; vulnary;
CC antiparasitic; antiparkinsonian; nootropic; neuroprotective; osteopathic;
CC anticonvulsant; antiarthritic; immunosuppressant; immunostimulant;
CC cardiant; thrombolytic; coagulant; vasotrophic; antidiabetic; hypotensive;
CC dermatological; immunosuppressive; antiinflammatory; antibacterial;
CC antiviral; antifungal; antirheumatic; antithyroid; and antianemic. The
CC sequences can be used for determining the presence of or predisposition
CC to, or preventing or treating pathological conditions associated with an
CC ORFX-associated disorder. The nucleic acids can be used to express ORFX
CC proteins in gene therapy vectors. The proteins and nucleic acids may be
CC used to treat cancers, proliferative disorders, neurodegenerative
CC disorders, osteoarthritis, graft vs host disease, cardiovascular disease,
CC diabetes mellitus, hypertension, hypothyroidism, cholesterol ester
CC storage, systemic lupus erythematosus, severe combined immunodeficiency
CC (SCID), AIDS, viral, bacterial or fungal infection, malaria, autoimmune
CC disorders, asthma, allergies, aplastic anaemia, burns, wounds, bone and
CC cartilage damage, nocturnal haemoglobinuria, antiinflammatory disease; to
CC enhance coagulation; to inhibit thrombosis; and as a contraceptive.
XX Sequence 130 AA;
SQ
Query Match 9.0%; Score 94.5; DB 21; Length 130;
Best Local Similarity 25.7%; Pred. No. 0.0038;
Matches 35; Conservative 22; Mismatches 58; Indels 21; Gaps 6;
QY 81 YLCVMGYQMLRGALKKEAVSAPAPQVELAKSGRSFLKGLTLNLANPKAIIYFGSVFLFV 140
DB 5 yllayatwrd--rsafamndtptvatars--llirgflnlnpkltiffiaflpqfv 60
QY 141 GDNVGTTA-----RMGIFALIIIVETLAWFTVVASLFALPQMRGY---QRLAKWIDGF 190
DB 61 --tpggtapaqlmivlsgvmamtlavfylvllanvf----rravvesprvqnllr 113
QY 191 AGALFAGFGIHLISR 206
DB 114 fatafaglglnlafaq 129
RESULT 11
R87527
ID R87527 standard; Protein; 153 AA.
XX R87527;
XX 01-JUL-1996 (first entry)
XX Mel-linked mlgA gene product.
XX Marine mela; selectable marker; oyster larva settlement;
KW pigmentation; melanin; p-hydroxyphenylpyruvate-hydroxylase;
KW mlgA gene.


```
XX 09-DEC-1988; 88DE-3841454.
XX PA
XX (DEGS ) DEGUSSA AG.
XX Kassing F, Kalinowski J, Arnold W, Winterfeldt A, Puhler A;
XX Kautz P-S, Thierbach G;
XX WPI; 1990-186465/25.
XX DR NPSDB; Q04775.
XX
XX Site specific mutagenesis of DNA at restriction enzyme positions -
XX by treating corresp. restriction fragments with hydroxylamine,
XX and new mutated sequences and transformed microorganisms
XX
XX Disclosure; ; p; German.
XX
XX When expressed in host cells this confers Chloramphenicol resistance
XX on them.
XX
XX Sequence 390 AA;

Query Match 8.2%; Score 86; DB 11; Length 390;
Best Local Similarity 20.6%; Pred. NO. 0.14;
Matches 47; Conservative 35; Mismatches 78; Indels 68; Gaps 10;

QY 1 MLMFLTAMVHIVALMSPGDPDFEFVSQ-----TAVSRSRKEAMMG 41
   ::: | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 71 lllvclllvfgshvlgamtvpflllitrslsalanagflavalstatlvpqkgrals 130
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 42 VL--GITCGVMVWAGI---ALLGLHLIIEKMAWLHTLMVGGGLYLCWGYQMLRGA--- 93
   :| | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 131 illsgttiatvv--gvpagallgtalgwrttfwaiaficipaav-----gvirgvtnn 181
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 94 LKKEAVSAPAPQVELAKSGRSFLKGLLTNLANPKAIIYFGSVSIFVGDNVGTTARMGIF 153
   : : | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 182 vgrsetsatsprlrv-----lsqatprlil----amalgalinggtfaaftl 227
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 154 ALIIVETL----AWFTVVASLFALPQMRGYSQRLAKWIDGFAGALFAG 197
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db 228 apivtetaglaawvsvvalvmfg-----lgsflgvtiaq 261
   | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

Search completed: May 6, 2001, 14:38:49
Job time: 395 sec

Db 67 MWRKGMANLNTSSQI-----SNRALITQGFVTAIANKPGWAFFMISLLPPFISVDQAIA 122
QY 148 ARWGIFALIIVETLAWFTVVA 168
Db 123 PQLMWLLSIIMMT-EFFSMLA 142

RESULT 2

5474933-4
; Patent No. 5474933
; APPLICANT: EINER, RONALD M.; FUQUA JR., WILLIAM C.
; TITLE OF INVENTION: MARINE MELA GENE
; NUMBER OF SEQUENCES: 9
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/148,945
; FILING DATE: 08-NOV-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 974,837
; FILING DATE: 10-NOV-1992
; APPLICATION NUMBER: 496,804
; FILING DATE: 21-MAR-1990
; SEQ ID NO: 4:
; LENGTH: 153
5474933-4

Query Match 8.6%; Score 91; DB 6; Length 153;
Best Local Similarity 25.5%; Pred. No. 0.0064;
Matches 36; Conservative 31; Mismatches 54; Indels 20; Gaps 6;

QY 43 LGYTCTGV--MWV-----AGIAL-----LGLHLIIKMAWLHTLIMVGGGLYLCWNGYQ 88
Db 7 LGMISGVRTLMWVGELAGVAIAAAMVGVASMLNYPQLFDILKMWGGGLYLGIGIS 66
QY 89 MLRGALKKEAVSAPQVELAKSGRSFL-KGLLTNLANKPAIIYFGSVFSLFVGDNVGTT 147
Db 67 MWRKGMANLNTSSQI-----SNRALITQGFVTAIANKPGWAFFMISLLPPFISVDQAIA 122
QY 148 ARWGIFALIIVETLAWFTVVA 168
Db 123 PQLMWLLSIIMMT-EFFSMLA 142

RESULT 3

US-08-311-023-2
; Sequence 2, Application US/08311023
; Patent No. 5693465
; GENERAL INFORMATION:
; APPLICANT: MANNING, David Lockwood
; APPLICANT: NICHOLSON, Robert Ian
; APPLICANT: GEE, Julia Margaret
; APPLICANT: GREEN, Christopher Douglas
; TITLE OF INVENTION: METHODS FOR PREDICTING THE BEHAVIOUR OF
; NUMBER OF SEQUENCES: 3
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Young & Thompson
; STREET: 745 South 23rd Street
; CITY: Arlington
; STATE: VA
; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.

; REGISTRATION NUMBER: Reg. No. 5693465 32.925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 431 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-311-023-2

Query Match 8.2%; Score 86.5; DB 1; Length 431;
Best Local Similarity 27.4%; Pred. No. 0.085;
Matches 49; Conservative 21; Mismatches 56; Indels 53; Gaps 11;
QY 53 AGIALGLHLIIKMAWLHTLIMVGGGLYLCWNGYQMLRGALKKEAVS----- 101
Db 263 AGVATL-----AW--WVINGDGLHNFSDGLAI--GAAFTEGLSSGLSTSVAVFCH 308
QY 102 PAPQ-----VELAKSGRSFLKGLLTNLANKPAIIYFGSVFSLFVGDNVGTTARWGIFALI 156
Db 309 ELPHELGDFAVLLKAGMTVQAVLYN-ALSAMLAYLGMATGIFIGHYAENVSMW-IFAL- 365
QY 157 IVETLAWFTVVASLFAFPOM-----RRGYORLAKWIDGFAGALFAGFIHLII 204
Db 366 ---TAGLFMYVALVDMVPEMLHNDASDHGCSRWGYFFLQN-----AGMLL-GFGIMLLI 415

RESULT 4

US-09-335-409-11
; Sequence 11, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 11
; LENGTH: 713
; TYPE: PRT
; ORGANISM: Sorangium cellulosum
US-09-335-409-11

Query Match 7.8%; Score 82; DB 4; Length 713;
Best Local Similarity 22.0%; Pred. No. 0.55;
Matches 46; Conservative 29; Mismatches 76; Indels 58; Gaps 9;

QY 1 MMLFLATVAMVHIVALMSGPGDPFFVTSQTAVSRSRKEAMGVL--GITCGVMWAGIAL 58
Db 9 VLLSLVTLALILVTAR-----ASGELARRLRQPEVLGELFGVGLGPSV-VGALAP 58
QY 59 GLH-----LIIEKMAWLHTLIMVGGGLYLCWNGYQMLRGALKKEAVSAPQVEL 108
Db 59 GFHRLAQEPAGVGVLSGISW-----IGALLLLMAGIEVDVGLRKEA----- 102
QY 109 AKSGRSFLKGLLTNLANKPAIIYFGSVFSLFVGDNVGTTARWGIFALIIVETLAWFTVVA 168
Db 103 ----RPGALSALGAIAADPLAA--GAAFSALVLDRLPS---GLFIGIVLSVTAYSVIAK 152
QY 169 SLFALPQMRRGY-----QRLAKWI 187

Db 960 PIVKSSLLYASSQALPFFCMALGFYWGSLGHGEYSLFQFYV 1002

RESULT

```

US-09-183-253-4
: Sequence 4, Application US/09183253
: Patent No. 6043054
: GENERAL INFORMATION:
: APPLICANT: VAWTER, LISA
: APPLICANT: STAMMERS, MELANIE
: TITLE OF INVENTION: NOVEL COMPOUNDS
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Katcher & Prestia
: STREET: P.O. Box 980
: CITY: Valley Forge
: STATE: PA
: COUNTRY: USA
: ZIP: 19482
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible
: OPERATING SYSTEM: DOS
: SOFTWARE: FastSEQ for Windows Ver
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/09/183,253
: FILING DATE: 30-OCT-1998
: CLASSIFICATION:
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 9817907.0
: FILING DATE: 17-AUG-1998
: APPLICATION NUMBER: 60/075,306
: FILING DATE: 20-FEB-1998
: ATTORNEY/AGENT INFORMATION:
: NAME: Prestia, Paul F
: REGISTRATION NUMBER: 23,031
: REFERENCE/DOCKET NUMBER: GP-70395
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 610-407-0700
: TELEFAX: 610-407-0700
: TELEX: 846169
: INFORMATION FOR SEQ ID NO: 4:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 332 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-09-183-253-4

```

```

Query Match          7.38; Score 77; DB 3; Length 332;
Best Local Similarity 21.74; Pred. No. 0.72;
Matches 50; Conservative 31; Mismatches 69; Indels 80; Gaps 11;

QY  2 LMFLTVAWVHIVALMSRGPDPFFVQSVARSRRK-----EAMGMVIGITCGVWVWAGTA 56
      | | : : : : : | : | : | : | : | : | : | : | : | : | : | : | : |
Db  96 LPYLSLSALTILGIMASAFLEF---NIKNRNOKLIKMSPYNNLIILGGMLSYASIF 152
      | | : : : : : | : | : | : | : | : | : | : | : | : | : | : | : |

QY  57 LGLL--HLIEK-----MAWLHT-----72
      | | : : : : : | : | : | : | : | : | : | : | : | : | : | : | : |

Db  153 LFLGDSGFVSEKTFETLCTVTRITLITGYTTFAGMFAKTRVHAIFKNVKKKKIKDQ 212
      | | : : : : : | : | : | : | : | : | : | : | : | : | : | : | : |

QY  73 --LIMVGGGL-----YLCWGMQYLMRGALKKEAVSAPAPQVELAKSGRSF-LKGLLTLNL 123
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db  213 KLEAVIVGMMLLDLCILICWQAVDPLRRTVEKYSME-PDP-----AGRDISIRPLLEHC 265
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

QY  124 ANPKALIEGVSFSLFVGDNYQTARTMGIFALIIVETLAVFTVVASLFAF 173
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |
Db  266 ENTHMTIWLGIIVYAY-----KGLMLFCGF-----LAWETRNVSPAL 303
      | | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : |

```

RESULT 9

US-06-612-734B-2
; Sequence 2, Application US/08612734B
; Patent No. 5914246
; GENERAL INFORMATION:

```
; APPLICANT: Peery, Robert B.
; APPLICANT: Skatrud, Paul L.
; APPLICANT: Tobin, Matthew B.
; TITLE OF INVENTION: Multiple Drug Resistance Gene of
; TITLE OF INVENTION: Aspergillus Fumigatus
; NUMBER OF SEQUENCES: 9
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center, DC1501
; CITY: Indianapolis
; STATE: Indiana
; COUNTRY: U.S.A.
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/612,734B
; FILING DATE: 08-MAR-1996
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Craig, Anne I.
; REGISTRATION NUMBER: 32,976
; REFERENCE/DOCKET NUMBER: X-9681
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 617-354-9570
; TELEFAX: 617-354-4043
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1349 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-612-734B-2

Query Match 7.3%; Score 76.5; DB 2; Length 1349;
Best Local Similarity 15.9%; Pred. No. 5.5;
Matches 37; Conservative 29; Mismatches 76; Indels 91; Gaps 5;

QY 3 MLETVAMVHVALMSPGDPFFVSVQTAVSRSRKEA----- 38
DB 826 LMFFVVGIAQFISLUSINGTAFATCERLRARRAKSQAFRSILRQDISFFPDREENSTGALTS 885
QY 39 -----MMGVGITCGVMVWAGIALGLHLILIEKMAWLHPLIMV----- 76
DB 886 FLSTETKNLSGSGVTGLTGMTSTTLGAAMIATAIGMKLALVCISVVPILLACGGLRF 945
QY 77 -----GGGLYLCHWGYQMLRGALKKEAVSAPAPQV-----E 107
DB 946 YMLAQFOQRKSAYEGSASTAC-----EATSAITVVASLTREQDVGVVYHDQ 992
QY 108 LAKSGRSLKGLLTN-----LANPKAIIYFGSVFLFVGDVNGVTARWGIFALII 157
DB 993 LOKGRKSLISVLRSLLYSQALVFCVAGLFWYGGTLLGHHEYSIFRFV 1045

RESULT 11
US-08-922-837-2
; Sequence 2, Application US/08922837
; Patent No. 5888770
; GENERAL INFORMATION:
; APPLICANT: Chalker,, Allison F.
; APPLICANT: Feliu,, Maria M.2.
; APPLICANT: Brown,, James R.
; APPLICANT: Bryant,, Alexander P.
; TITLE OF INVENTION: No. 5888770el SpoilIE
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: dechert, Price & Rhoads
; STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
```

```
; CITY: Philadelphia
; STATE: PA
; COUNTRY: USA
; ZIP: 19103-2793
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette
; COMPUTER: IBM Compatible
; OPERATING SYSTEM: DOS
; SOFTWARE: FastSeq for Windows Version 2.0
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/922,837
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Dickinson, Q. Todd
; REGISTRATION NUMBER: 28,354
; REFERENCE/DOCKET NUMBER: GM10077
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 215/994-2252
; TELEFAX: 215/994-2222
; TELEX:
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 783 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
US-08-922-837-2

Query Match 7.0%; Score 73.5; DB 2; Length 783;
Best Local Similarity 25.3%; Pred. No. 5.8;
Matches 49; Conservative 24; Mismatches 64; Indels 57; Gaps 12;

QY 35 RKEAM-----MGVLGITCGVMV-----WAGIALGLHLILIEKMAW 69
DB 36 RKEAQRMLISIGIAILLIFAAFKLGAAGITLYNLRLVGLSVLAIVLGLLIYLFEEK 95
QY 70 -----LHTLMVGGGLYLCHWGYQMLRGALKKEAVSAPAPQVELAKSG---RSFLKGL 119
DB 96 IRKQGLLSGFFTFAGLLIFAYLVWKYGLDKSVLKGTMQAVVDTLGTFTTSFAGG 155
QY 120 LTNLANPKAIIYFGSVFLFVGDVNGVTARWGIFALIIIVETL---AWFTV-VASLFPALPQ 175
DB 156 LIGVA-----LYIPTAF-LF--SNIGTYFIGST--LILVGLLVSPWSVYDIAEFS--- 202
QY 176 MRRGYQRLAKWIDG 189
DB 203 --RGP---AKWEG 211

RESULT 12
US-08-750-723A-2
; Sequence 2, Application US/08750723A
; Patent No. 5981219
; GENERAL INFORMATION:
; APPLICANT: Flugge, Ulif-Ingo
; APPLICANT: Weber, Andreas
; APPLICANT: Fischer, Karsten
; TITLE OF INVENTION: DNA MOLECULES WHICH CODE FOR A PLASTID
; TITLE OF INVENTION: 2-OXOGLUTARATE/MALATE TRANSLOCATOR
; NUMBER OF SEQUENCES: 2
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: James F. Haley, Jr., Esq c/o FISH & NEAVE
; STREET: 1251 Avenue Of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10020
; COMPUTER READABLE FORM:
```

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent In Release #1.0, Version #1.30 (EPO)
CURRENT APPLICATION DATA: US/08/750,723A
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: DE P 44 20 782.4
FILING DATE: 15-JUN-1994
ATTORNEY/AGENT INFORMATION:
NAME: James F. Haley, Jr., Esq.
REGISTRATION NUMBER: 27,794
REFERENCE/DOCKET NUMBER: GFB-3
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 596-9000
TELEFAX: (212) 596-9090
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 569 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-750-723A-2

Query Match 6.9%; Score 73; DB 2; Length 569;
Best Local Similarity 20.8%; Pred. No. 4.3;
Matches 43; Conservative 24; Mismatches 56; Indels 84; Gaps 8;

QY 34 SRKEAMGV-LGITGVVMWAG-----TALLGLHLII-----EKMAW--- 69
DB 362 TRNESTMAVTLTGLVFGKLGVDVTAAILGLSVLLITGVVTKCLAESVAMDYL 421
QY 70 -----LHFLIMVGGGLYLCWMGYQMLRGAL----- 94
DB 422 TWFAALIAMAGLYNKYGLITFSENVKVVVGGGLGSW---QMSFGVLVLLYFYSHYFFAS 478
QY 95 -----KKEAVSAPAPQVELAKSGRSFLKGLLTNLANPKAIITFGSVSLFV 140
DB 479 GRAHIGAMPTAFLVSASALGTPPFLAAIVLSFLSNLMLGLTHYIGSAPVFGANY---- 534
QY 141 GDNVGTARWGI-FALIIVETLAWTV 166
DB 535 ---VPLPQWVGFLISIVNLIWLGV 558

RESULT 13
US-08-924-747-8
Sequence 8, Application US/08924747
Patent No. 6083570
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
TITLE OF INVENTION: ENZYMES
NUMBER OF SEQUENCES: 32
CORRESPONDENCE ADDRESS:
ADDRESSER: E.I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: UNITED STATES OF AMERICA
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/924,747
FILING DATE:

CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: FLOYD, LINDA AXAMETHY
REGISTRATION NUMBER: 33,692
REFERENCE/DOCKET NUMBER: CL-1108
TELECOMMUNICATION INFORMATION:
TELEPHONE: 302-892-8112
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 234 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: not relevant
MOLECULE TYPE: protein
ORIGINAL SOURCE: SOYBEAN
TISSUE TYPE: SOYBEAN
IMMEDIATE SOURCE:
CLONE: SE3.03B09
US-08-924-747-8

Query Match 6.7%; Score 71; DB 3; Length 234;
Best Local Similarity 25.0%; Pred. No. 2.2;
Matches 29; Conservative 15; Mismatches 48; Indels 24; Gaps 6;

QY 84 WMGY-----QMLRGALKKEAVSAPAPQVELAKSGRSFLKGLLTNLANPKAIITFGSVFS 137
DB 100 WFAYIDEKWFVTSLSVLVAEDDEAKKPHFEQAEGLERLEEVEFNKYSEGA--YFG---- 153
QY 138 LFVGDNVG-TTARWGI-FALIIVETLAWTVVWASLFAIPOM-RRGYORLAKWIDGFA 191
DB 154 ---GDSIGFIDIGFSF-----LSWMRVIEEMSGRKLDEKKHPGLTOWAETFA 199

RESULT 14
US-09-247-373B-8
Sequence 8, Application US/09247373B
Patent No. 6168954
GENERAL INFORMATION:
APPLICANT: MCGONIGLE, BRIAN
APPLICANT: O'KEEFE, DANIEL
TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE ENZYMES
FILE REFERENCE: CL-1108-A
CURRENT APPLICATION NUMBER: US/09/247,373B
CURRENT FILING DATE: 1999-02-10
PRIOR APPLICATION NUMBER: 08/924,747
PRIOR FILING DATE: 1997-09-05
NUMBER OF SEQ ID NOS: 56
SOFTWARE: Microsoft Office 97
SEQ ID NO 8
LENGTH: 234
TYPE: PRT
ORGANISM: SOYBEAN
US-09-247-373B-8

Query Match 6.7%; Score 71; DB 4; Length 234;
Best Local Similarity 25.0%; Pred. No. 2.2;
Matches 29; Conservative 15; Mismatches 48; Indels 24; Gaps 6;

QY 84 WMGY-----QMLRGALKKEAVSAPAPQVELAKSGRSFLKGLLTNLANPKAIITFGSVFS 137
DB 100 WFAYIDEKWFVTSLSVLVAEDDEAKKPHFEQAEGLERLEEVEFNKYSEGA--YFG---- 153
QY 138 LFVGDNVG-TTARWGI-FALIIVETLAWTVVWASLFAIPOM-RRGYORLAKWIDGFA 191
DB 154 ---GDSIGFIDIGFSF-----LSWMRVIEEMSGRKLDEKKHPGLTOWAETFA 199

RESULT 15
US-09-296-715-8
Sequence 8, Application US/09296715

```

; Patent No. 6171839
; GENERAL INFORMATION:
; APPLICANT: MCGONIGLE, BRIAN
; APPLICANT: O'KEEFE, DANIEL
; TITLE OF INVENTION: SOYBEAN GLUTATHIONE-S-TRANSFERASE
; NUMBER OF SEQUENCES: 32
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: E.I. DU PONT DE NEMOURS AND COMPANY
; STREET: 1007 MARKET STREET
; CITY: WILMINGTON
; STATE: DELAWARE
; COUNTRY: UNITED STATES OF AMERICA
; ZIP: 19898
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.50 INCH
; COMPUTER: IBM PC COMPATIBLE
; OPERATING SYSTEM: MICROSOFT WORD FOR WINDOWS 95
; SOFTWARE: MICROSOFT WORD VERSION 7.0A
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/296,715
; FILING DATE:
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: FLOYD, LINDA AXAMETHY
; REGISTRATION NUMBER: 33,692
; REFERENCE/DOCKET NUMBER: CL-1108
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 302-892-8112
; TELEFAX: 302-773-0164
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 234 amino acids
; TYPE: amino acid
; STRANDEDNESS: not relevant
; TOPOLOGY: not relevant
; MOLECULE TYPE: protein
; ORIGINAL SOURCE:
; TISSUE TYPE: SOYBEAN
; IMMEDIATE SOURCE:
; CLONE: SE3.03B09
; US-09-296-715-8

```

```

Query Match      6.7%; Score 71; DB 4; Length 234;
Best Local Similarity 25.0%; Pred. No. 2.2;
Matches 29; Conservative 15; Mismatches 48; Indels 24; Gaps 6:

QY   84 WMGY-----QMLRGALKKEAVSAPAPQVELAKSGRSFLKGLLTNLNLPKAIYFGSVFS 137
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   100 WPAYIDEKWFTSLRSVLVAEDDEAKKPHFEQAEGLERLEEVFNKYSEGKA--YFG--- 153
      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY   138 LFVGDNVG-TTARWGIFALIIVEILAWFTVVASLFALPQM-RRGYORLAKWIDGFA 191
      ||::|| | | | | | | | | | | | | | | | | | | | | | | | | | |
Db   154 ---GDSIGFDIDFGGSF-----LSMMRVITEMSGRKLLDEKKHPLGTOWAETFA 199

```

Search completed: May 6, 2001, 14:39:24
Job time: 415 sec

OM of: US-09-466-935-4 to: GenEmbl.* out_format : pfs

Date: May 6, 2001 3:43 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL-frame1.p2n.model -DEV-x10  
-D/cgn2_1/USF10.spool/US09466935/runat_03052001_075709_14798/app_query.fasta_1.529  
-QB-GenEmbl -QFMT=fastap -Suffix=rge -GAPOP=12.000 -GAPEXT=4.000  
-MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.500 -GAPOP=4.500  
-QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000  
-FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000  
-DELEXT=7.000 -START=1 -MATRIX=blomsun62 -TRANS=human4.0.cdi  
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0  
-ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0  
-MAXLEN=200000000 -USER=US09466935 @CGNL_1.5396 -NCPU=6  
-ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1
```

Search information block:

```
Query: US-09-466-935-4  
Query length: 206  
Database: GenEmbl.*  
Database sequences: 1283235  
Database length: -1216004940  
Search time (sec): 1745.950000
```

score_list:

Sequence	Strd	Orig	ZScore	EScore	Len	Documentation
gb_pat1:AX030177	+	1054.00	1898.96	1.5e-97	840	! AX030177 Sequence 3 from Patent
gb_bal:AE0005613	+	1054.00	1874.21	3.5e-96	10592	! AE0005613 Escherichia coli O15
gb_bal:AE000458	+	1027.00	1824.50	2.1e-93	11509	! AE000458 Escherichia coli K12
gb_bal:3:STCOW85	+	998.00	1751.74	2.3e-89	91414	! M87049 E. coli genomic sequen
gb_bal:3:BYGTM01	+	983.00	1724.08	8.1e-88	96086	! AF233324 Salmonella typhimur
gb_bal:AC073768	-	954.00	1661.13	2.6e-84	279589	! AC073768 Mus musculus clone
gb_bal:AC020970	+	747.00	1288.88	1.4e-63	212936	! AC020970 Mus musculus clone
gb_bal:AC020885	-	717.00	1232.29	2.0e-60	268294	! AC020885 Mus musculus clone
gb_bal:AC079167	+	559.00	945.36	1.9e-44	290452	! AC079167 Mus musculus chromo
gb_bal:2:ECOREC0	+	486.00	858.85	1.3e-39	2695	! M30198 E.coli recQ gene comple
gb_bal:AC020893	-	411.00	692.35	2.4e-30	62274	! AC020893 Mus musculus clone K
gb_bal:AE006110	+	370.50	636.68	3.0e-27	10184	! AE006110 Pasteurella multocid
gb_bal:032810	+	368.50	632.19	5.4e-27	11137	! U32810 Haemophilus influenzae
gb_bal:AE004109	-	313.50	532.41	1.9e-21	11321	! AE004109 Vibrio cholerae chr
gb_bal:AP001517	+	292.00	461.66	1.7e-17	294250	! AP001517 Bacillus halodurans
gb_pat1:AX063725	+	283.00	503.15	8.3e-20	792	! AX063725 Sequence 7 from Patent
gb_bal:AE004719	+	274.50	462.85	1.5e-17	10145	! AE004719 Pseudomonas aerugin
gb_bal:3:YEFUABC	+	269.50	460.86	1.9e-17	4923	! Z47200 Versinia enterocolitica
gb_bal:AE004589	+	263.50	440.55	2.5e-16	12948	! AE004589 Pseudomonas aerugin
gb_pat1:AX030079	+	261.50	465.81	9.9e-18	6722	! AX030079 Sequence 9 from Patent
gb_bal:AE000140	-	261.50	437.27	3.9e-16	12498	! AE000140 Escherichia coli K12
gb_bal:AE005212	-	258.50	433.98	5.9e-16	10029	! AE005212 Escherichia coli O15
gb_bal:AP003009	-	255.00	393.24	1.1e-13	339681	! AP003009 Mesorhizobium loti
gb_bal:AF235020	-	251.50	425.93	1.7e-15	6246	! AF235020 Brucella melitensis s
gb_bal:BCU73857	-	251.50	396.37	7.3e-14	128824	! U73857 Escherichia coli chrc
gb_bal:AE004699	+	246.50	411.87	1.0e-14	10426	! AE004699 Pseudomonas aerugin
gb_bal:AE004864	-	245.00	408.65	1.5e-14	10976	! AE004864 Pseudomonas aerugin
gb_bal:AC087482	+	244.00	398.33	5.7e-14	26245	! AF087482 Pseudomonas aerugin
gb_bal:AC022157	+	217.50	366.86	3.2e-12	4833	! AC022157 Homo sapiens chromos
gb_bal:AF270347	-	214.50	367.20	3.1e-12	2629	! AF270347 Staphylococcus epide
gb_bal:AF269817	-	214.50	366.31	3.5e-12	2575	! AF269817 Staphylococcus epide
gb_bal:AE004946	-	208.50	340.84	9.1e-11	13075	! AE004946 Pseudomonas aerugin
gb_bal:AE004889	+	207.00	339.45	1.1e-10	11415	! AE004889 Pseudomonas aerugin
gb_bal:AP001508	+	201.00	296.76	2.6e-08	296950	! AP001508 Bacillus halodurans
gb_bal:AP003001	+	193.50	292.55	4.4e-08	345783	! AP003001 Mesorhizobium loti
gb_bal:3:ACARAB	+	190.00	339.86	1.0e-10	2483	! U04992 Pseudomonas aeruginosa
gb_bal:PAU81259	+	193.00	329.34	4.0e-10	7285	! U81259 Pseudomonas aeruginosa
gb_bal:AP004412	-	197.50	318.46	1.6e-09	16814	! AE004412 Vibrio cholerae chr
gb_bal:AP002995	-	197.00	327.99	8.0e-08	346897	! AP002995 Mesorhizobium loti
gb_pat1:AX030081	+	190.00	336.81	1.5e-10	639	! AX030081 Sequence 11 from Patent
gb_bal:AE000274	-	190.00	306.81	7.1e-09	13793	! AE000274 Escherichia coli K12

```
gb_bal:D90824 - 190.00 306.50 7.4e-09 14236 ! D90824 E.coli genomic DNA,  
gb_bal:D90823 - 190.00 305.69 8.2e-09 15475 ! D90823 E.coli genomic DNA,  
gb_bal:AF188935 + 190.00 287.84 8.1e-08 96231 ! AF188935 Bacillus anthraci  
gb_bal:AP001519 - 186.00 269.37 8.7e-07 303650 ! AP001519 Bacillus halodur  
seq_name: gb_pat1:AX030177
```

seq_documentation_block:

```
LOCUS AX030177 840 bp DNA PAT 16-SEP-2000  
DEFINITION Sequence 3 from Patent EP1013765.  
ACCESSION AX030177  
VERSION AX030177.1 GI:10190394  
KEYWORDS  
SOURCE  
ORGANISM Escherichia coli.  
Escherichia coli.  
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;  
Escherichia.  
1 (bases 1 to 840)  
Balareva,A.V., Aleshin,V.V., Livshits,V.A., Tokhmakova,I.L. and  
Zakataeva,N.P.  
Gene and method for producing l-amino acids  
Patent: EP 1013765-A 3 28-JUN-2000;  
AJINOMOTO KK (JP)  
FEATURES  
source  
1..840  
/organism="Escherichia coli"  
/db_xref="taxon:562"  
187..807  
/note="unnamed protein product"  
/codon_start=1  
/transl_table=11  
/protein_id="CAC09319.1"  
/db_xref="GI:10190395"
```

CDS

alignment_scores:

```
Quality: 1054.00 Length: 206  
Ratio: 5.117 Gaps: 0  
Percent Similarity: 100.000 Percent Identity: 100.000
```

alignment_block:

```
US-09-466-935-4 x AX030177 ..  
Align seg 1/1 to: AX030177 from: 1 to: 840
```

```
1 MetLeuMetLeuPheLeuThrValAlaMetValHisTleValAlaLeuMe 17  
187 AGTTGTAGTATTCTTCCCGTCCGATGGTGACATTGTGGCGCTTAT 236  
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArg 34  
237 GAGCCCGGTCGCCGATTCTTTTGTCTCTCAGACGCGCTGCAGTCGTT 286  
34 erArgLysGluAlaMetMetGlyValLeuGlyThrCysGlyValMet 50  
287 CCCGTAAAGACGATGATGGCGTGTGGCATACCTCGCGCGTAATG 336  
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisIleIleIleIle 67  
337 GTTGGGCGGGGATGGCGCTCTGGCTGCTGATGATCGAAAAAT 386  
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyLeuTyrLeuCys 84  
387 GGCGTGGCTGCATACGCTGATGATGGCGGGGCGGCTGTATCTCTGCT 436  
84 rpMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
```

```
437 GGATGGCTTACCAGATCTACGTGGTGACCTGAAAGAGCGGTTTCT 486
101 AlaProAlaProGlnValGluLeuAlaIySSerGlyArgSerPheLeuIy 117
487 GCACCTCGCCACAGGTGCGAGCTGGCGAAAGTGGCGCAGTTTCTGAA 536
117 sGlyLeuLeuThrAsnLeuAlaasnProIySAlaIleIleIyPheGlys 134
537 AGGTTTACTGACCAATCTCGCTAATCCGAAGCGATTATCTACTTTGGCT 586
134 erValPheSerLeuPheValGlyAspAsnValGlyThrAlaArgTrp 150
587 CGGTTCTCATTTGTCGCTGATACGTTGGCATCTACCGCGCGTGG 636
151 GlyTlePheAlaLeuIleIleValGluThrLeuAlaTrpPheThrValVa 167
637 GGCATTTTTCGCTGATCATCTTCGAAACGCTGGCGTGTATACCGTCGT 686
167 lAlaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuA 184
687 TGCAGCGCTTTTGGCCCTCGCCCAATGCGCGTGGTTATCAACGTCGG 736
184 lAlySTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyTle 200
737 CGAAGTGGATTGATGTTTGGCGGCGGCTTATTTGGCCGATTGGGCATT 786
201 HisLeuIleIleSerArg 206
787 CATTTGATTATTTCGCGG 804

seq_name: gb_bai:AE005613

seq_documentation_block:
LOCUS AE005613 10592 bp DNA BCT 25-JAN-2001
DEFINITION Escherichia coli O157:H7 genome, contig 3 of 3, section 232 of 290.
ACCESSION AE005613 AE005174
VERSION AE005613.1 GI:12518685
KEYWORDS
SOURCE Escherichia coli O157:H7.
ORGANISM Escherichia coli O157:H7
Bacteria: Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.
1 (bases 1 to 10592)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,D.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Genome sequence of enterohemorrhagic Escherichia coli O157:H7
Nature 409 (6819), 529-533 (2001)
2 (bases 1 to 10592)
Perna,N.T., Plunkett,G. III, Burland,V., Mau,B., Glasner,J.D.,
Rose,D.J., Mayhew,G.F., Evans,P.S., Gregor,J., Kirkpatrick,H.A.,
Postai,G., Hackett,J., Klink,S., Boutin,A., Shao,Y., Miller,D.,
Grotbeck,E.J., Davis,N.W., Lim,A., Dimalanta,E., Potamousis,K.,
Apodaca,J., Anantharaman,T.S., Lin,J., Yen,G., Schwartz,D.C.,
Welch,R.A. and Blattner,F.R.
Direct Submission
Submitted (22-OCT-2000) Laboratory of Genetics, University of
Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
FEATURES
Source
1..10592
Location/Qualifiers
/organism="Escherichia coli O157:H7"
/strain="EDL933"
/serotype="O157:H7"
/db_xref="taxon:83334"
/note="enterohemorrhagic"
<1..71
misc_feature
/note="O-island #155; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(72..836)
/gene="z5332"

gene
complement(72..836)
/gene="z5332"
/function="orf; Unknown function"
/note="Residues 1 to 131 of 254 are 95.41 pct identical to
residues 1 to 131 of 161 from Escherichia coli K-12 Strain
MG1655; B3815"
/codon_start=1
/protein_id="AAG59008.1"
/db_xref="GI:12518686"
/translation="MAHRLIGKGMITLNLKRIFLALTLPLFAVAADDCALSDPTLT
VOATVNPQTERVKMVKQKANGAWGTFLHALLADMSQGVOMAMNGGIYDSYAPLG
LYIENCOKVALNLASGEGNFFRPGGVFYAGKVGIVLDAFKTSKEIOPAVQSGP
MLMENSVINPRIHENVASRKIRGCVGINKHGNAYFLLSQATNFYDFACYAKALUNE
QLLYLDGTSISHMYMKGGAIPWQRIYFPVIMISVERKG"
1206..2156
/gene="corA"
/note="z5333"
1206..2156
/gene="corA"
/function="transport; Transport of small molecules:
cations"
/note="Residues 1 to 316 of 316 are 100.00 pct identical
to residues 1 to 316 of 316 from Escherichia coli K-12
Strain MG1655; B3816"
/codon_start=1
/transl_table=11
/product="Mg2+ transport, system I"
/protein_id="AAG59009.1"
/db_xref="GI:12518687"
/translation="MLSAFOLENNRLTRLEVESQPLVNAVWIDLVDPDDDERLRVQS
ELGSLATRPELEDIEASARFEDDGLHIHSFFFEDEADHAGNSTVATIRDCGLR
PIRELPAPFLYBMBARSQSMVDGNAYELLDFETKLEQADEIENLYSDLEQLSR
VIMEGQGDVEDALSTLALEDIGWVRLCLMDTORALNFLVRKARLPGGLEQARE
IURDESLLPNESILFQKVNFLMQAMGFINEQNRIRKIFSVSVVFLPFLVASSY
GMNFEPMLKWSFGYGAIFMILAGLAPLYFKRKNML"
2157..4848
/note="O-island #156; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(2198..2689)
/gene="z5334"
complement(2198..2689)
/gene="z5334"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59010.1"
/db_xref="GI:12518688"
/translation="MKSLOKILCLFLCCIYNLWLRPVOILTYSDAGNSVFLVV
DHLPTDSDKINWYLKHQNEIKNOHPLGEGSWHTVWVIDNGFTDYKKYIEGYPEDL
YCFPTIKSNDNCITKNYLMVINEYPYRNTGHINDFTYQLTQENKIEQVFNPFKK
DNF"
complement(2686..3534)
/gene="z5335"
complement(2686..3534)
/gene="z5335"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59011.1"
/db_xref="GI:12518689"
/translation="MNNYLLPSINIPCTLETISLFEDNYSAADMOYGDVVEODFLSLG
LSDIAKVDYPRLLKRYHPPGGSINVAFSASSGSKIISQRECTDILFAEMKELAMFS
FPGQYKTLIEDLIIEHFYRNGSNFHSQOLNFKHINKYINSPIRNIKEICENGIN
STPGYQPLIQISKTKLLSSRLNKNFEDFSGNLGIVSHDIAQRIISLISFOKYA
IGWSATIEFVAQDHFGLDVTDIKNKTSKYRFRFRIFWFLQHKDFAFKPFPTNTIE
RIENYL"
3748..3867
/gene="z5336"
3748..3867
/gene="z5336"

misc_feature
complement(72..836)
/gene="z5332"

gene
complement(72..836)
/gene="z5332"
/function="orf; Unknown function"
/note="Residues 1 to 131 of 254 are 95.41 pct identical to
residues 1 to 131 of 161 from Escherichia coli K-12 Strain
MG1655; B3815"
/codon_start=1
/protein_id="AAG59008.1"
/db_xref="GI:12518686"
/translation="MAHRLIGKGMITLNLKRIFLALTLPLFAVAADDCALSDPTLT
VOATVNPQTERVKMVKQKANGAWGTFLHALLADMSQGVOMAMNGGIYDSYAPLG
LYIENCOKVALNLASGEGNFFRPGGVFYAGKVGIVLDAFKTSKEIOPAVQSGP
MLMENSVINPRIHENVASRKIRGCVGINKHGNAYFLLSQATNFYDFACYAKALUNE
QLLYLDGTSISHMYMKGGAIPWQRIYFPVIMISVERKG"
1206..2156
/gene="corA"
/note="z5333"
1206..2156
/gene="corA"
/function="transport; Transport of small molecules:
cations"
/note="Residues 1 to 316 of 316 are 100.00 pct identical
to residues 1 to 316 of 316 from Escherichia coli K-12
Strain MG1655; B3816"
/codon_start=1
/transl_table=11
/product="Mg2+ transport, system I"
/protein_id="AAG59009.1"
/db_xref="GI:12518687"
/translation="MLSAFOLENNRLTRLEVESQPLVNAVWIDLVDPDDDERLRVQS
ELGSLATRPELEDIEASARFEDDGLHIHSFFFEDEADHAGNSTVATIRDCGLR
PIRELPAPFLYBMBARSQSMVDGNAYELLDFETKLEQADEIENLYSDLEQLSR
VIMEGQGDVEDALSTLALEDIGWVRLCLMDTORALNFLVRKARLPGGLEQARE
IURDESLLPNESILFQKVNFLMQAMGFINEQNRIRKIFSVSVVFLPFLVASSY
GMNFEPMLKWSFGYGAIFMILAGLAPLYFKRKNML"
2157..4848
/note="O-island #156; Region of the EDL933 chromosome not
homologous to E. coli K-12 MG1655"
complement(2198..2689)
/gene="z5334"
complement(2198..2689)
/gene="z5334"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59010.1"
/db_xref="GI:12518688"
/translation="MKSLOKILCLFLCCIYNLWLRPVOILTYSDAGNSVFLVV
DHLPTDSDKINWYLKHQNEIKNOHPLGEGSWHTVWVIDNGFTDYKKYIEGYPEDL
YCFPTIKSNDNCITKNYLMVINEYPYRNTGHINDFTYQLTQENKIEQVFNPFKK
DNF"
complement(2686..3534)
/gene="z5335"
complement(2686..3534)
/gene="z5335"
/function="orf; Unknown function"
/note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59011.1"
/db_xref="GI:12518689"
/translation="MNNYLLPSINIPCTLETISLFEDNYSAADMOYGDVVEODFLSLG
LSDIAKVDYPRLLKRYHPPGGSINVAFSASSGSKIISQRECTDILFAEMKELAMFS
FPGQYKTLIEDLIIEHFYRNGSNFHSQOLNFKHINKYINSPIRNIKEICENGIN
STPGYQPLIQISKTKLLSSRLNKNFEDFSGNLGIVSHDIAQRIISLISFOKYA
IGWSATIEFVAQDHFGLDVTDIKNKTSKYRFRFRIFWFLQHKDFAFKPFPTNTIE
RIENYL"
3748..3867
/gene="z5336"
3748..3867
/gene="z5336"
```

```
/function="orf; Unknown function"
/Note="Residues 1 to 39 of 39 are 100.00 pct identical to
residues 278 to 316 of 316 from GenPept 118 :
gi|6960250|gb|AAE33440.1| (AF233324) S. typhimurium
magnesium transport protein (CORA) (SP:P31138) [Salmonella
typhimurium LT2]"
/codon_start=1
/transl_table=11
/product="partial repeat of corA"
/protein_id="AAG59012.1"
/db_xref="GI:12518690"
/translation="MNFEPPELKWFSGYPGAIIEMLAGUAPLYFKRKNWL"
/complement(3892..4278)
/gene="25337"
/complement(3892..4278)
/gene="25337"
/function="orf; Unknown function"
/Note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59013.1"
/db_xref="GI:12518691"
/translation="MFKSSIRNIFPYSKLERLSRFQFIYILTWSILITNDINLAL
SFILVINTVYNINIGAKRIRDTGCKAIYFVLISLGLTLFSIAAIEGVRCESLSVQ
LNVAWILFTIMLSAIPSAKMSLSAH"
/gene="25339"
/483..4794
/gene="25339"
/function="orf; Unknown function"
/Note="No significant matches"
/codon_start=1
/transl_table=11
/protein_id="AAG59014.1"
/db_xref="GI:12518692"
/translation="MDSIWEMVTFGSCSWKNPKYSKSDPVKAWLLWAIITLVFIL
SLTTLNLRSLIGLSLGFQCIILGRYTMSLGSWKYMGVLGGPPVTNIVLILIFPKK
R"
complement(4841..5728)
/gene="25340"
/complement(4841..5728)
/gene="25340"
/function="orf; Unknown function"
/Note="Residues 4 to 295 of 295 are 92.12 pct identical to
residues 8 to 299 of 300 from Escherichia coli K-12 Strain
MG1655: B3819"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59015.1"
/db_xref="GI:12518693"
/translation="MDAKQTRQGVLLAALAAFYFWGTAPEFKLIYVPADELILHRVI
WSFFWVMSICQNSYLKTLQTKIFMLAVSAILGGNWLFIWAVNHHLEA
SLGFNPLNVLIGMLIGERFRMQWLAVILACGLVQLWTFGSPILALGAFS
FAPGLVRRKIAVEAQGMLETWMLLPVAAIYLFADISDSHMGPMSPMLNLLIA
AGIVTPLLCTFAATRLRLSLTGLGFFQYIGPTLMFLAVTFYGEKPGADRMVTFPI
WVALAFVMDAIYTORRS"
complement(5780..6265)
/gene="25341"
/complement(5780..6265)
/gene="25341"
/function="orf; Unknown function"
/Note="Residues 1 to 161 of 161 are 100.00 pct identical
to residues 1 to 161 of 161 from Escherichia coli K-12
Strain MG1655: B3820"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAG59016.1"
/db_xref="GI:12518694"
/translation="MKCKDDMSAVLTAEQALKLVGMFVYHMPFNRLGMELERYEKE
```

```
FAQLAFKNQPMVGNNAQSILHGGVITASALDVAAGLVCVGSTLIRHETISEDELQRRL
SRMTGIDLRVDYLPRGXERFTATSSLLRAGNKVAVARVELHNEQLYIASATATYKV
G"
6412..7281
/gene="plda"
/Note="25342"
6412..7281
/gene="plda"
/function="enzyme; Central intermediary metabolism:
phosphorus compounds"
/Note="Residues 1 to 289 of 289 are 100.00 pct identical
to residues 1 to 289 of 289 from Escherichia coli K-12
Strain MG1655: B3821"
/codon_start=1
alignment_scores:
Quality: 1054.00 Length: 206
Ratio: 5.117 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-466-935-4 x AE005613
Align seg 1/1 to: AE005613 from: 1 to: 10592
1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
9307 ATGTTGATGTTATTTCTCACCGTCGCCATGTGCACATTGTGGCGCTTAT 9356
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
9357 GAGTCCCGGTCCCGATTTCTTTTGTCTCTCAGACCGCTGTCAGTCGTT 9406
34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
9407 CCCGTAAGAAGCGATGATGGCGTCTGGGTATTACCTGGCGGCTTATG 9456
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMe 67
9457 GTTTGGCAGGATGCGTGTGCTGGCTGCATTTGATTATCGAAAAAT 9506
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyLeuTyrLeuCyst 84
9507 GCCTGTGCTGCATACGCTGATATGTTGGCGGTGGCTGTATCTCTGCT 9556
84 rpMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
9557 GGATGGGTTACACATGCTGCGCGCGCTGAAAAAGAGCGCGTTCT 9606
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLy 117
9607 GCACCTGGCCACAGGTCGAGCTGGCGAAGAGTGGCGCAGTTTCTGAA 9656
117 sGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleIleThrPheGlyS 134
9657 AGGTTACTGACCAATCTCGCTAATCCGAAAGCGATATTACTTTGGCT 9706
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTirp 150
9707 CGGTGTTCTCATTTGTTGCGGTGATAAGCTTGGCACTACCGCGCGCTGG 9756
151 GlyIlePheAlaLeuIleIleValGluThrLeuAlaTrpPheThrValVa 167
9757 GGCATTTTCGGTGTGATTAATGTCGAAACGCTGGCGTGTGTTACTGTGTT 9806
167 lAlaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuA 184
9807 GCGCAGCTGTTTGGCCCTGCCGCAATGCGCGCTGTTATCAACGCTCTGG 9856
184 lAlaLysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
9857 CCAAGTGGATTGATGTTTTCGCGGGCGGTATTTCGCGGATTTGGCATT 9906
```

```

201 HisieurlelleSerArg 206
|||||
9907 CATTGATTATTCGCGG 9924

seq_name: gb_ba1:AE000458

seq_documentation_block:
LOCUS      AE000458      11509 bp      DNA      BCT      01-DEC-2000
DEFINITION Escherichia coli K12 MG1655 section 348 of 400 of the complete
genome.
ACCESSION AE000458 U00096
VERSION   AE000458.1 GI:2367299
KEYWORDS .
SOURCE   Escherichia coli K12.
ORGANISM Escherichia coli K12.
          Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
          Escherichia.
REFERENCE 1 (bases 1 to 11509)
          Blattner, F.R., Plunkett, G. III, Bloch, C.A., Perna, N.T., Burland, V.,
          Riley, M., Collado-Vides, J., Glasner, J.D., Rode, C.K., Mayhew, G.F.,
          Gregor, J., Davis, N.W., Kirkpatrick, H.A., Goeden, M.A., Rose, D.J.,
          Mau, B., and Shao, F.
          The complete genome sequence of Escherichia coli K-12
          Science 277 (5331), 1453-1474 (1997)
          97426617
REFERENCE 2 (bases 1 to 11509)
          Blattner, F.R.
          Direct Submission
          Submitted (16-JAN-1997) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
REFERENCE 3 (bases 1 to 11509)
          Blattner, F.R.
          Direct Submission
          Submitted (02-SEP-1997) Guy Plunkett III, Laboratory of Genetics,
          University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
          Email: ecoligenetics.wisc.edu Phone: 608-262-2534 Fax:
          608-263-7459
REFERENCE 4 (bases 1 to 11509)
          Plunkett, G. III
          Direct Submission
          Submitted (13-OCT-1998) Laboratory of Genetics, University of
          Wisconsin, 445 Henry Mall, Madison, WI 53706, USA
          On Sep 9, 1997 this sequence version replaced gi:1790254.
          This sequence was determined by the E. coli Genome Project at the
          University of Wisconsin-Madison (Frederick R. Blattner, director).
          Supported by NIH grants HG00301 and HG01428 (from the Human Genome
          Project and NCHGR). The entire sequence was independently
          determined from E. coli K12 strain MG1655. Predicted open reading
          frames were determined using GeneMark software, kindly supplied by
          Mark Borodovsky, Georgia Institute of Technology, Atlanta, GA.
          30332 [email: mark@amber.gatech.edu]. Open reading frames that
          have been correlated with genetic loci are being annotated with CG
          Site Nos., unique ID nos. for the genes in the E. coli Genetic
          Stock Center (CGSC) database at Yale University, kindly supplied by
          Mary Berlyn. A public version of the database is accessible
          (http://cgsc.biology.yale.edu). Annotation of the genome is an
          ongoing task whose goal is to make the genome sequence more useful
          by correlating it with other data. Comments to the authors are
          appreciated. Updated information will be available at the E. coli
          Genome Project's World Wide Web site
          (http://www.genetics.wisc.edu). *** The E. coli K12 sequence and
          its annotations are periodically updated; this is version M54. No
          sequence changes. Annotation updates: updated gene identifications
          and products; all new functional assignments courtesy of Monica
          Riley; added promoters, protein binding sites, and repeated
          sequences described in reference 1. The unique numeric identifiers
          beginning with a lowercase 'b' assigned to each gene (protein- or
          RNA-encoding) are now designated as gene synonyms instead of
          labels. This should allow them to be searched for in Entrez as gene
          names.

FEATURES
         source
Location/Qualifiers
1..11509
/organism="Escherichia coli K12"
/strain="K12"
/sub_strain="MG1655"
/db_xref="taxon:83333"
81..950
/gene="plda"
/note="b3821"
81..950
/gene="plda"
/EC_number="3.1.1.32"
/function="enzyme; Central intermediary metabolism:
Phosphorus compounds"
/note="o289; 99 pct identical to PAL_ECOLI SW: P00631;
CGSCNo. 384"
/codon_start=1
/transl_table=11
/product="outer membrane phospholipase A"
/protein_id="AAC76824.1"
/db_xref="GI:2367300"
/translacion="MRTLQGNLLPVMPLPMVYQAENVKEVHDPAVYRGSIANNMLO
EHDNPFLLPYDITNYLIYTQSLNKEATASYNDEPQFLGATDYRFAQWTDLRVEMGT
GPNVLGASYTQKSWQLSNSESSPFRTNYPQLFGLGATDYRFAQWTDLRVEMGT
NHDNNGRSQDTSRSWNRLYTRLMAENGNNLVVEKPMYVVGNTDDNPTTKTMYGYYLQK
ICYHLGDVAVLSAKQYNNWNTGYGGAEGLSLYPTTKHVRLYTQVYSGYGESLIDYNFNO
TRVGVGVMINDLF"
965..993
/note="factor Sigma70; predicted +1 start at 4003392"
1019..1047
/note="factor Sigma70; predicted +1 start at 4003446"
1077..2909
/gene="recQ"
/note="b3822"
1077..2909
/gene="recQ"
/EC_number="3.6.1.-"
/function="enzyme; DNA - replication, repair,
restriction/modification"
/note="o610; 99 pct identical to 607 amino acids
of RECO_ECOLI SW: P15043 but has 3 additional N-ter
residues"
/codon_start=1
/transl_table=11
/product="ATP-dependent DNA helicase"
/protein_id="AAC76825.1"
/db_xref="GI:2367301"
/translacion="MNVQAQEVNLNLSGAKQVLQETFGYQQFRPQGEIIDTVLSGRD
CLVTWPTGGGKSLCYQIPALLNGLTVVYVSPILSMKQVQDQOLQANGVAAACLNSTQ
REQOLEVMTGCGTQIRLLYIAPERLMLDNLEHLAHWNPNVLLAYDEAHCISQNGHDF
RPEALQGLQRORFPTLPFMALTATADDTTQDIDVRLGLNDPLLIQISFORPNIRYM
LMERKPLDOLMRVYQQRGSGIIYCNSRAKVEDTAARLQSGISAAAYHAGLENNV
RADYQEKFRDDLIQVVAQVAFGNGINKPNRVFVHFDIPRNIESYQYQETGRACRDGL
PAEAMLYPDADMALPRCLLEKPOGLODIERHKLNAMGFAEAQTCRRLLVLYRG
EGRQEPGNCIDICLDPPKQDGTGDAQIALSIGVNGQFGYGVVEVIRGANNQRI
DYGHDKLVYGMGRKDSHERWSVIRQLIHGLVTONIAQHSALQLTAEARPYLAESS
LQLAVPRIVALKPRAMOKRSGNVDKRLFAKLKRLKLSIADESNNVPYVVFVNDATLIE
MAEQMPTIASEMLSYNGVGMKLERFGKPFMALIRAHVGDGDEE"
2104..2140
/note="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
3162..3189
/note="factor Sigma70; predicted +1 start at 4005588"
3224..3592
/gene="yigU"
/note="b3823"
3224..3592
/gene="yigU"
/function="orf; Unknown"
/note="o122; sequence change shortens and
changesN-terminus relative to earlier version (YIG_ECOLI
SW: P27846)"
/codon_start=1
repeat_region
promoter
gene
CDS

```

```

/transl_table=11
/product="orf, hypothetical protein"
/db_xref="GI:2367302"
/translation="MGYQMLRGALKKEAVSAPQVVELAKSGRSLKGLLTNLNLPKA
IIFYGSVFSFLFVGNVGTARWGIFALLIIVETLAWPTVVVASLFPQWRGYYQLAKW
IDGAFAGLAFAGFIHLISR"
repeat_region
3612..3645
/feature="REP (repetitive extragenic palindromic) element;
contains 1 REP sequence"
complement(3654..4070)
/gene="yigK"
/feature="b3824"
complement(3654..4070)
/gene="yigK"
/feature="yigK"
/function="orf; Unknown"
/feature="f138; 100 pct identical to yigK_ECOLI SW: P27847"
/codon_start=1
/transl_table=11
/product="orf, hypothetical protein"
/protein_id="AAC76826.1"
/db_xref="GI:1790258"
/translation="MIAFEVLKWAGAYLIWLIQWRAAGALDKSLASTQSRHFL
QRAVFNLTNPKSIVFLAALFPQFIMPQPOLMQLYVLGVTTIVVDIIVMIGYATLQ
RIALWIKGPKOMKALNKFISLSFLYGALLASARHA"
4200..4230
/feature="factor Sigma70; predicted +1 start at 4006629"
complement(4227..4254)
/feature="factor Sigma70; predicted +1 start at 4006612"
4384..5406
/gene="p1dB"
/feature="b3825"
4384..5406
/gene="p1dB"
/EC_number="3.1.1.5"
/function="enzyme; Central intermediary metabolism;
Phosphorus compounds"
/feature="o340; 99 pct identical amino acid sequence and
equal length to P1DB_ECOLI SW: P07000"
/codon_start=1
/transl_table=11
/product="lysophospholipase L(2)"
/protein_id="AAC76828.1"
/db_xref="GI:2367303"
/translation="MFOQKQDWETRENAFAFTMGPLTDFWRQDEAEFTGVDDIPVR
FVFRQAHDVVVICPGRIEYVKYAEALYDLHLGFDVLIIDHRGSGRGLLADP
HLGHVNRNDYVDDLAAPWQOEVPQGPWRKRYILAHSMGCAISTLFLQHRPGVCDAIA
LTAPMFGIVIRMPSEMAROILNWAEPHPRDGYAIGTGRWRALPAINVLTHSRQRY
RNLRFYADPTIRVGPTVHWVRESILAGEOVLGAGODATPTLLIQAEFEFVYDNR
HMDRCLELRTAAGHPVGGRLPIKGYHEILFEKDMASVALHAIIVDFNRHNSPFG
NKSTEV
5381..5408
/feature="factor Sigma70; predicted +1 start at 4007807"
5426..5455
/feature="factor Sigma70; predicted +1 start at 4007854"
5438..5466
/feature="factor Sigma70; predicted +1 start at 4007865"
5494..5523
/feature="factor Sigma70; predicted +1 start at 4007922"
5526..5641
/gene="yigL"

```

alignment_scores:

Quality: 1027.00 Length: 206
Ratio: 5.010 Gaps: 0
Percent Similarity: 99.515 Percent Identity: 99.515

alignment_block:

US-09-466-935-4 x AE000458 ..

Align seg 1/1 to: AE000458 from: 1 to: 11509

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17

```

|||||
2973 ATGTTGATGTTATTTCTACCGTCGCCATGGTGACATTTGTCGCCCTTAT 3022
17 tserProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
|||||
3023 GAGCCCGGTCGCCATTTCTTTTGTCTCAGACCGCTGTCAGTCGTT 3072
34 arArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
|||||
3073 CCCGTAAAGAGCATGATGGCGTCTGGGCATTACCTGGCGCGTAATG 3122
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleGluLysMe 67
|||||
3123 GTTTGGGCTGGGATTCGCTTGGCCTGCATTGATTATCGAAAAAT 3172
67 talatrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyrLeuCyst 84
|||||
3173 GGCCTGGCTGCATACGCTGATTATGGTGGCGGTGGCTG.TATCTCTGCT 3221
84 rpMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
|||||
3222 GATGGGTTACCATGCTACGTGGTGCACCTGAAAAAGAGCGGTTTCT 3271
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLys 117
|||||
3272 GCACCTGGCGCACAGTCGAGCTGGCGAAAAGTGGCGCAGTTTCTGAA 3321
117 sgLysLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGlys 134
|||||
3322 AGGTTACTGACCAATCTCGCTAATCCGAAAGCGATTATCTACTTGGCT 3371
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrp 150
|||||
3372 CGGTGTTCTCATTTGTTCGGTGATAACGTTGGCATTACCGCGCGCTGG 3421
151 GlyIlePheAlaLeuIleIleValGluThrLeuAlaTrpPheThrValVa 167
|||||
3422 GGCATTTTGGCGTGATCATTCGAAACGCTGGCGTGGTTTACCGTCTGT 3471
167 llaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuA 184
|||||
3472 TCCAGCGCTGTTGCCCTGGCGAATGCCCGCTGGTTATCAACGCTCTGG 3521
184 laLysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
|||||
3522 CGAAGTGGATGATGGTTTGGCGGGCGGTATTTCGCGGATTGGCATT 3571
201 HisLeuIleIleSerArg 206
|||||
3572 CATTCGATTATTCGCGG 3589

```

seq_name: gb_ba2.ECOUW85

seq_documentation_block:

LOCUS ECOUW85 91414 bp DNA BCT 29-MAY-1995
DEFINITION E. coli genomic sequence of the region from 84.5 to 86.5 minutes.
ACCESSION M87049
VERSION M87049.1 GI:836656
KEYWORDS
SOURCE Escherichia coli.
ORGANISM Escherichia coli
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Escherichia.

REFERENCE 1 (bases 1 to 91414)

AUTHORS Daniels,D.L., Plunkett,G. III., Burland,V. and Blattner,F.R.

TITLE Analysis of the Escherichia coli genome: DNA sequence of the region

from 84.5 to 86.5 minutes

JOURNAL Science 257 (5071), 771-778 (1992)

MEDLINE 92358234

REFERENCE 2 (bases 1 to 91414)

AUTHORS Plunkett,G. III., Burland,V., Daniels,D.L. and Blattner,F.R.

TITLE Analysis of the Escherichia coli genome. III. DNA sequence of the

region from 87.2 to 89.2 minutes

JOURNAL Nucleic Acids Res. 21 (15), 3391-3398 (1993)
 MEDLINE 93347969
 REFERENCE 3 (bases 1 to 91414)
 AUTHORS Blattner,F.R., Burland,V., Plunkett,G. III., Sofia,H.J. and Daniels,D.L.
 TITLE Analysis of the Escherichia coli genome. IV. DNA sequence of the region from 89.2 to 92.8 minutes
 JOURNAL Nucleic Acids Res. 21 (23), 5408-5417 (1993)
 MEDLINE 94089392
 REFERENCE 4 (bases 1 to 91414)
 AUTHORS Rudd,K.E., Sofia,H.J., Koonin,E.V., Plunkett,G. 3rd, Lazar,S. and Rouviere,P.E.
 TITLE A new family of peptidyl-prolyl isomerases
 JOURNAL Trends Biochem. Sci. 20 (1), 12-14 (1995)
 MEDLINE 95184296
 REFERENCE 5 (bases 1 to 91414)
 AUTHORS Daniels,D.L.
 TITLE Direct Submission
 JOURNAL Submitted (12-FEB-1992) Donna L. Daniels, Laboratory of Genetics, University of Wisconsin, 445 Henry Mall, Madison, WI 53706, USA.
 MEDLINE Email: ecoligenetics.wisc.edu; Phone: 608-262-2534; Fax: 608-263-7459
 REFERENCE On May 29, 1995 this sequence version replaced gi:148169.
 JOURNAL This sequence was determined as part of the E. coli Genome Project (Frederick R. Blattner, director) at the University of Wisconsin-Madison. The entire sequence was independently determined from E. coli MG1665. Overlaps and conflicts with other sequence determinations are annotated. Reference [1] describes the original sequence determination of a 91408 bp sequence. References [2], [3], and [4] describe subsequent corrections and/or updates to that sequence.

FEATURES Location/Qualifiers
 Source 1..91414 "Escherichia coli"
 /organism="Escherichia coli"
 /strain="K-12"
 /sub_strain="MG1655"
 /db_xref="taxon:562"
 /clone_lib="lambda library of Daniels and Blattner"
 /map="bp 1 at 3974 kb; 84.7 minutes"
 /note="This sequence comprises the following lambda clones: DD850(EC15-116), DD851(EC13RM4), DD854(EC15-177), DD856(EC19RM8.1), DD858(EC17RM9), DD864(EC21-76), DD861(EC17-221), DD864(EC27-149), DD865(EC21-76), DD867(EC27-860). M13mp19 or Janus vectors were used for subcloning."
 misc_feature 1..200
 /note="corresponds to K02846 (1..200)"
 misc_feature 1..5863
 /note="clone EC15-116"
 misc_feature complement(1..4899)
 /note="corresponds to M37337 (1..4900)"
 misc_feature 1..339
 /note="corresponds to V00349 (164..500)"
 tRNA 10..86
 /gene="aspT"
 /note="CGSC No. 989"
 gene 10..86
 /gene="aspT"
 tRNA 95..170
 /gene="trpT"
 /note="CGSC No. 66"
 gene 95..170
 /gene="trpT"
 terminator 177..205
 /note="terminator-like sequence"
 misc_structure 253..396
 /note="predicted bend of 81 degrees"
 gene complement(266..862)
 /gene="f198"
 CDS complement(266..862)
 /gene="f198"
 /codon_start=1
 /transl_table=11

misc_structure 586..724
 /function="unknown"
 /protein_id="AAA67566.1"
 /db_xref="GI:148170"
 /translation="MAHTSRNFEISGASLSMECMNLNWLGRLYNQDAHTGLOPEA
 RIAQSVLSKQHLHERGDLITTEAPKMDFFSSQLGYFTLALYTSAPSKLGGDNLIL
 RLEWMPDQHQEHLIGADEVPIITTSAAEQQAQIAMLNGCTWLEFVSWARKKGLHT
 VVDSTLSRLPYAIWLQNSDKNALIRDLKINVLDEY"
 /note="predicted bend of 77 degrees"
 complement(703..1104)
 /gene="f133"
 complement(703..1104)
 /gene="f133"
 /note="CGSC No. 18010; possible pssR gene"
 /codon_start=1
 /transl_table=11
 /protein_id="AAA67567.1"
 /db_xref="GI:148171"
 /translation="MDTELLKTFLEVSRTFRHFGRAAESLYLTQSAYSFRIQLENQLG
 VNLFRHNRNIRLTAAAGEKLLPYAETLMTWQAARKRWRIPHDITSLFSLVPAPRCGNV
 CLISGDACIKIRMPIQAYSKRELPLNGSLM"
 complement(1136..1165)
 /note="promoter matrix score of 63"
 1151..1179
 /note="promoter matrix score of 57"
 1223..11636
 /gene="o137"
 1223..11636
 /gene="o137"
 /codon_start=1
 /transl_table=11
 /function="unknown"
 /protein_id="AAA67568.1"
 /db_xref="GI:148172"
 /translation="MAESFTTNRYFNKHYPGFSRHGDFTIKEAQLLHGHYAFNE
 LDLGKREPVEEKLFAVCGEREPEVTEARVMSKYMTRIKRPRFHTLSAVNRRLK
 VLKTPILTINKGEMPFYVSSFCRWMSRSIAR"
 1562..11586
 /gene="o137"
 /note="terminator-like sequence"
 complement(1585..3135)
 /gene="f516"
 complement(1585..3135)
 /gene="f516"
 /note="similar to Mg chelatase subunit of Rhodobacter capsulata"
 /codon_start=1
 /transl_table=11
 /function="unknown"
 /protein_id="AAA67569.1"
 /db_xref="GI:148173"
 /translation="MHIMVTQGGILMSLSIVHTRAAALGVNAPPTIVFVHLSKGLPGLTM
 VGLPETVKEARDVRNSAINSGVEYPAKKITINLAPADLPKEGGYDLPALIALAA
 SEQLPANKLDEYELVGEALTGALRGVPGAISATKSGRIIVAKDNEDEGLIN
 GEGCLTADHLQAVCAFLEGHKHALEXPKPTDVARLOHDLSDVIGQEQKRGLETTAA
 GGHNLLIGPFGTKTMLASINGLLPDLSNEEALESAAILSLVNAESVQKQMRQPF
 RSPHSLATAMVGGGAIPTGFEISLAHNGVLELDELPEFERTLDALREPTESQIIF
 LSRTAKITYPARQLVAAMNPSPTGHYQGNHNRCTPEQTLRYLNLRLSGPFLDRPDL
 LEIPLPPGILSKTVVPGESSATVKQVMAARERQFKRONKLNALDLSPEIRQFCKLE
 SEDAMWLEGTLLHLGLSIRAWQLLKVARTIADQSDITITRQHLQEAHSVYRAIDRL
 THLOKLT"
 3287..3433
 /note="predicted bend of 79 degrees"
 3297..3326
 /note="promoter matrix score of 74"
 3336..3676
 /note="corresponds to M11651 (1..341)"
 3336..3737
 /note="corresponds to V00289 (1..402)"
 3336..12791
 /note="corresponds to M10313 (1..9456)"
 3379..5721
 /note="V00290; ECILYX(1..2343)"

```
gene
3458..3556
/gene="ilvL"
CDS
3458..3556
/gene="ilvL"
/codon_start=1
/transl_table=11
/product="ilvGEDA operon leader peptide"
/protein_id="AAA67570.1"
/db_xref="GI:148174"
/translation="MFALLRVISLVISVVVLIIPPCGAALGRGA"
3696..4679
/gene="ilvG"
3696..4679
/gene="ilvG"
/feature="CGSC No. 603; inactive in wild-type E. coli K12"
/codon_start=1
/transl_table=11
/product="acetylhydroy acid synthase II, large subunit"
/protein_id="AAA67571.1"
/db_xref="GI:148175"
/translation="MNGAQVYVHALRAQGVNTVFGPGGAIMPVYDALYDGGVEHLIC
RHEGQAMAAIGYARATGKTGVCIATSGPGATNLITGLADALLDSIPVVAITGQVSAP
FIGTADQEVFLGLSLACTKHSFLVQSLSEELPRIMAEAFVACSGRCPVLVDIPKD
IQLASGDLEPFTVEVNEVAFHAEVEQAKLAKQKPMLYVGGVGMAQVAPALRE
FLAATKMFATCTLKLGLGAEADYPYILGMLGMHGTKAANFAVQEGDILLIATGAREDDR
VTGKLNTFAPHASVTHMDIDPAEMNKLRLQAHVALQGLDNLALLPALQQLNQ"
4675..5340
/gene="o221"
4675..5340
/gene="o221"
/feature="o221"
/feature="the mutation ilvO2096 joins this ORF to ilvG,
restoring function missing in wild-type"
/pseudo
/codon_start=1
/variation
replace(4677..4678,"tatg")

alignment_scores:
Quality: 998.00 Length: 206
Ratio: 4.892 Gaps: 0
Percent Similarity: 99.029 Percent Identity: 99.029

alignment_block:
US-09-466-935-4 x ECOW85 ..
Align seg 1/1 to: ECOW85 from: 1 to: 91414
1 MetLeuMetLeuThrValAlaMetValHisIleValAlaLeuMe 17
60863 AAGTTGATGTTATTCACCGTCGCCATGGTGACATTTGGCGCTTAT 60912
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
60913 GAGCCCGCGTCCCGATTCTCTTTTCTCTCAGACCGCTGTCTAGTCT 60962
34 erArGlyGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
60963 CCCGTAAAGAGCGATGATGGCGTGTGGCATTTACCTGGCGCGTATG 61012
51 ValTrpAlaGlyIleAlaLeuGlyLeuHisLeuIleGluLysMe 67
61013 GTTTGGGCTGGGATGGCGTGTGGCGTGCATTTGATTATCGAAAAAT 61062
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyrlLeuCyst 84
61063 GGCCTGGCTGCATACGCTGATTATGGTGGCGGTGGCTG.TATCTCTGCT 61111
84 rpMetGlyTyrlGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
61112 GGATGGGTACCAGATCTACGTGGTGCATGNAAGAGAGCGGTTCTCT 61161
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLy 117
61162 GCACCTGGCCACAGGTCGAGCTGGCGAAAGTGGCGCAGTTCTCTGAA 61211
```

```
117 sGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleIleTyrPheGlyS 134
61212 AGT.TTACTGACCAATCTCGCTAATCCGAAGCGATTATCTACTTTGGCT 61260
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrrp 150
61261 CGGTGTTCTCATTTGTTCGGTGAATACGTTGGCCTACCGCGCGCTGG 61310
151 GlyIlePheAlaLeuIleIleValGluThrLeuAlaIlePheThrValVa 167
61311 GGCATTTTTCGCTGATCATTTGCGAACGCTGGCGTGGTTACCGCTGT 61360
167 lAlaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuA 184
61361 TGCAGCGCTGTTTTCGCTGCGCAATGCGCGTGGTTATCAACGCTGG 61410
184 lAlaYsTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
61411 CGAAGTGATGATGTTTTCGCGGCGCTTATTTGCCGATTGGCATT 61460
201 HisLeuIleIleSerArg 206
61461 CATTTGATTATTTGCGCG 61478
seq_name: gb_ba3:STVSTMD1
seq_documentation_block:
LOCUS STVSTMD1 DNA BCT 11-FEB-2000
DEFINITION Salmonella typhimurium fragment STMD1.
ACCESSION AF233324
VERSION AF233324.1 GI:6960215
KEYWORDS
SOURCE Salmonella typhimurium LT2.
ORGANISM Salmonella typhimurium LT2.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
Salmonella.
REFERENCE
1 (bases 1 to 96086)
AUTHORS Washington University Genome Sequencing Center.
TITLE The Salmonella typhimurium Genome Sequencing Project
JOURNAL Unpublished
REFERENCE
2 (bases 1 to 96086)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (09-FEB-2000) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
COMMENT
Submitted by:
Genome Sequencing Center
Department of Genetics, Washington University,
St. Louis, MO 63110, USA
e-mail: sclifton@watson.wustl.edu or
jspliethe@watson.wustl.edu
This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry or covered by high quality data (i.e., phred quality
>= 30); an attempt was made to resolve all sequencing problems,
such as compressions and repeats; all regions were covered by
sequence from more than one ml3 subclone.
NOTES:
Coding sequences below are predicted from manually evaluated
computer analysis, using similarity information and the programs
GLIMMER (Salzberg, S., Delcher, A., Kasif, A. and White, O. (1998)
NAR 26,544-548), and Genemark (Lukashin, A.V. and Borodovsky, M.
(1998), NAR 26,1107-1115.
Location/Qualifiers
1..96086
/organism="Salmonella typhimurium LT2"
/strain="SGSC1412"
FEATURES
source
```

```
/db_xref="taxon:99287"
/clone="STMD1"
<!. .980
/note="rric"
/product="23S ribosomal RNA"
1072. .1191
/note="rric"
/product="5S ribosomal RNA"
1244. .1323
/product="tRNA-Asp"
1329. .1401
/product="tRNA-Trp"
complement(1506. .2354)
/gene="STMD1.99"
complement(1506. .2354)
/gene="STMD1.99"
/note="83% identity with E. coli hypothetical protein
(YIFA) (SP:P22788) and 95% identity with amino acids 1-80
of E. coli possible regulatory protein (PSSR) (SP:P27826)"
/codon_start=1
/transl_table=11
/evidence=not_experimental
/protein_id="AAF33488.1"
/db_xref="GI:6960298"
/translation="MESTVDTLLKTFLEVSRTRHPCRAAEALYLTQSAVSFRIRQLE
NQLGVNFTRRNNIRLTAGEKLLPYAETLMNTQAAKEVAHTSRNEFSIGASIS
LWECMLNAGRLYQLQPOSGQLQFEARQAQSIVKQLHERQLDLITTTPAKMDEF
SSQLIGHTLALYCSPPARKSEINTLRLEWGFQFOHETGLIAADEVPVLTSSAEL
ARQOLSALNGCSWLPVNWANEKGLHTVADSATLSRPLVAIWLQNSDKYSLICDLTKT
DVLDEQ"
2461. .2799
/gene="yifB"
2461. .2799
/gene="yifB"
/codon_start=1
/transl_table=11
/product="95% identity with E. coli hypothetical protein
(YIFE) (SP:P27827)"
/protein_id="AAF33485.1"
/db_xref="GI:6960295"
/translation="MAESFTTNRYPDNKHPGFSRHGDFIKEAQLLRHGHAFND
LDLGRPEVTEELFVAVCGEREPEVTAERVWSKYMTRIKPRFHTLSGGRQVE
GAEDYTEADD"
complement(2824. .4344)
/gene="yifB"
complement(2824. .4344)
/gene="yifB"
/note="STMD1.95"
/codon_start=1
/transl_table=11
/product="85% identity to E. coli hypothetical protein
(YIPB) (SP:P22787); contains similarity to Pfam family
PF01078 (Magnesium chelatase, subunit Chl1), score=351.8,
E=7.6e-102, N=1"
/db_xref="GI:6960294"
/translation="MSLAIVHTRAALGVNAPPTIEVHISNGLPLTIVWGLPETVKE
ARDVRSAIINSGYEFPKKTITINLAPADLPKEGGYDLPPIAVALLAASEOLTASNLE
AYELVGLALTGALRGVPGAISATTAIRAGRNIVATENAAEVLISKEGCTIADHL
QYVCAFLGHALERPLAODMASPTATADLRDVIQEGCKRGLEITAAAGHNLILIGP
PGTGTWLASRLSGILPPLSNEALPESAAITLSLVNADTVKPMOORPFRSPHSASLT
AMVGGGAIAPGELSIAHNGILFDLPEFERTLDALREPIESQIHLSTRAKITIY
PARQLIAAPGSPSTGHQNNRCPTQETLRNLNLSGFPDLDREDFLSLEIPLPPGI
LSQHAGSSSATVKKRVIAAHERQRRKLNALREGLREIQYCVLHHDDARWLEDT
LVHGLSTRAWQRLKKVARTIADIELAQISRQHLQEAVSVAIRADRLLIHLQKLLA"
4696. .4794
/gene="ilvI"
4696. .4794
/gene="ilvI"
/note="STMD1.92"
/codon_start=1
/transl_table=11
```

```
/product="100% identity to E. coli ilvGMEDA operon leader
peptide (ILVL) (SP:P03060)"
/protein_id="AAF33486.1"
/db_xref="GI:6960296"
/translation="MTALLRVLISLVISVVVVIIPPCGAALGRGA"
4934. .6580
/gene="ilvG_1"
4934. .6580
/gene="ilvG_1"
/note="STMD1.91"
/codon_start=1
/transl_table=11
/product="92% identity with E. coli acetolactate synthase
II (ILVG) (SP:P00892) ; contains similarity to Pfam
family PF02005 (Thiamine pyrophosphate enzymes),
score=952.6, E=1.5e-295, N=1"
/protein_id="AAF33483.1"
/db_xref="GI:6960293"
/translation="MNGAQMWHALRAQGVKTVFGYPGGAIMPVYDALYDGVVEHLIC
RHEGQAMAAIGYARSTGKTGVCITATSGGATNLITGLADALLDSVPVVAITGOVSAP
FIGTDAQGEVDVLGSLACTKHSFLVQSLEELPRIMAEAFVANRGCPVLVDIPKD
IOLASGELEPMFTYANEATFPQADVEQAROMLEQAKPMLYGGVGAQAVPALRK
FIATQMFVCTILKGLGAEADIPYTLGLHGHGTRKANPAVQECDDLIIAVGAREDDR
VTGLNTPAPNASVHMDIDPAEMNLRQAHVALQDLSLPLAQQLPKLIDAMRQSC
AELRAEAWRMDGFEETIYAPLLKOLSERKPADSVYTVVDSQHQMWAQHWITRPE
NFIITSSGLTGMFGFLPAAGVAQVARDNVICISGDGSPMNVQELGVTKRQDPLKI
VLLDNRLGVWQMOQLFFQERYSETTLTNDPDLMLASAFIPGQHITRDKQVEAAL
DTMLASGPLYLLHVSIDELNWPVPPGASNSEMLEKLS"
6580. .6840
/gene="ilvM"
6580. .6840
/gene="ilvM"
/note="STMD1.90"
/codon_start=1
/transl_table=11
/product="93% identity with E. coli acetolactate synthase
II (ILVM) (SP:P13048)"
/protein_id="AAF33482.1"
/db_xref="GI:6960292"
/translation="MQHVNVSARFNPETLRLVRVVRHGFQVCMNMEAATDAQNI
NIETVASPRSDLLFLSQLKLVDAHVAICQSAATSQOIRA"
6858. .7787
/gene="ilvE"
6858. .7787
/gene="ilvE"
/note="STMD1.89"
/codon_start=1
/transl_table=11
/product="S. typhimurium branched-chain-amino-acid
transaminase (ILVE) (SP:P15168)"
/protein_id="AAF33481.1"
/db_xref="GI:6960291"
/translation="MTTKADYIFNCEMVRWEDAKVHVNHALHVGTSVREGIRCYD
SHKGVFRHREHMQRLDSAKYIRFPVQSIDELMEACRDVIRKNNTSAYIRPLVF
VDVGVMGPNPPGYTDTVIIAAPPWAGIAGAEALDQGDIDAMVSNWRAAPNTIAPAAK
AGNYLSSLLVGSSEARRHGVEQGLDANGYISEGAGNLFVEKDGVLFTPPFTSSAL
PCITRDALIKLAKELGIEVREQSLRESLYLADEVEMSGTAAEITFVRSVDGIQVGE
RCGPVTKRIQQAFFGLTGETDKWCWLDPVNS"
7948. .9798
/gene="ilvD"
7948. .9798
/gene="ilvD"
/note="STMD1.88"
/codon_start=1
/transl_table=11
/product="S. typhimurium dihydroxy-acid dehydratase (ILVD)
(SP:P40810); contains similarity to Pfam family PF00920
(dehydratase family), score=1023.9, E=3.5e-304, N=1"
/protein_id="AAF33480.1"
/db_xref="GI:6960290"
/translation="MPKYRSATTHGRNMAGARALWRATGMDTSDFGKPIIAVNSFT
QFVCHVHLRLGLKLVAEQIEASGVAKENTIAVDGGLAGMGLYSLPSRELIA
SVEVNVNHCADAMVCISNCDKITPCMLMASLRLNIPVIFVSGGFWAEGTKLSDKII
```


KLDLYDAMTLOGADPKVSDOSNOVERSNACPTGSCSCSMFTANSNCLTEALGLSQPGN
GSLIATHADRKOLFNLAKGRIVELTKRYEONDESALPRNIAKAAAFERNAMTLDIWMG
GSTNTVHLHLLAAQAEDIDFTMSDIDLKSRKVPOLCKVAPSTQKRYHMEVDVHRAGGVLG
ILGELDRALLNRNKNVLYGTLPOTLEQYDITVTOEAVKMFRAKMPAGIRRTQAFS
QCRWSDSDRAACIGRSLRYAYSKDGLAVLYNGFAENGICIVKTAGVDDSDILKFTG
PAKVYESQDDVEALIGKVVGVGVVIRYEGPKGPGMGQEMLYPTSFLLKSMGLKAC
ALITDGRFSGGTSGLSIGHVSPEAASGCTIALIEDGPTIADIPNRSIOLQISEABIA
ARREAOEARGDKAWTPKPNRQKQVSEFALRAYSLATSADKCAVRDKSKLGG
9801..11345

alignment_scores:
Quality: 983.00 Length: 206
Ratio: 4.842 Gaps: 0
Percent Similarity: 98.544 Percent Identity: 90.777

alignment_block:

US-09-466-935-4 x STYSTMD1 ..
Align seg 1/1 to: STYSTMD1 from: 1 to: 96086

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
60428 ATGATGATGCTATTTTACCGTGGCAATGGTGCACATCGTTGCGCTAAT 60477
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
60478 GAGCCCCGCGCCGACTTTTCTTCGTGTCTCAAACTGCTGTCAGCGGT 60527
34 erArgLysGluAlaMetMetClyValLeuGlyIleThrCysGlyValMet 50
60528 CCGGTAAAGAGCGCATGATGGCGTACTGGGCGATCACCTGCGCGGTGATG 60577
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGlyLysMe 67
60578 GTATGGCGGGCGTGGCGCTGTTGGCTGCATCTCATCATCAAAAAAAT 60627
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyLeuCysT 84
60628 GGCTTGGCTGCATACGATCATATGATGGCGGGCGGCTGTACCTGTGCT 60677
84 rPMeGlyTyTrGlnMetLeuArgGlyAlaLeuLysLysGlyAlaValSer 100
60678 GGATGGGTATCAGATGCTGCGCGCGCTGTAACAAAAACAGGATGCGGCG 60727
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuL 117
60728 GCSTCTCTCCGACATGTAACCTGCGCAGCGGGCGCAGCTTCTCAA 60777
117 scGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyPheGlyS 134
60778 AGGGCTGTGACCAATCTGTCGAATCTCTAAGCGGATATCATATTTGGTT 60827
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrp 150
60828 CCGTTTTTTCATTTTTCGCGCATAAAGCTGCGCGCTGGCGGCGCTCGG 60877
151 GlyIlePheAlaLeuIleValIleValIleValIleValIleValIleVal 167
60878 GGTATTTTCGGCTTAATCCCTCGAAGCGCTGCCCTGGTTTACCGTAGT 60927
167 lAlaSerLeuPheAlaLeuProGlnMetArgGlyTyTrGlnArgLeuA 184
60928 CGCCAGCGCTGTTCGCGCTCGCGAAATGCGCGCTGCGTATCATCGCTCGG 60977
184 lAlaYsrPheAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
60978 CGAATGATGATGATGGCTTCCCGCGCGCTCTGTTTGGCGGCTTTGTATT 61027
201 HisLeuIleIleSerArg 206
61028 CACCTGATATTTCGCGC 61045

seq_name: gb_htg16:AC073768

seq_documentation_block:
LOCUS AC073768 279589 bp DNA HTG 29-JUN-2000
DEFINITION Mus musculus clone RP23-359N18, WORKING DRAFT SEQUENCE, 60
unordered pieces.
ACCESSION AC073768
VERSION AC073768.1 GI:8810385
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE 1 (bases 1 to 279589)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 279589)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (29-JUN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov

Project Information
Center Project Name: 1875058
Center clone name: RPCI-23_359N18

Summary Statistics
Consensus quality: 230709 bases at least Q40
Consensus quality: 253950 bases at least Q30
Consensus quality: 258916 bases at least Q20
Estimated insert size: 216000; agarose-fp estimation
Estimated insert size: 273689; sum-of-contigs estimation
Quality coverage: 9.54 in Q20 bases; agarose-fp estimation
Quality coverage: 7.53 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 60 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1009: contig of 1009 bp in length
* 1010: gap of unknown length
* 1110: contig of 1023 bp in length
* 2133: gap of unknown length
* 2232: contig of 1025 bp in length
* 3257: gap of unknown length
* 3357: gap of unknown length
* 4550: contig of 1193 bp in length
* 4650: gap of unknown length
* 6414: contig of 1764 bp in length
* 6514: gap of unknown length
* 6515: contig of 1043 bp in length
* 7557: gap of unknown length
* 7657: contig of 1228 bp in length
* 8885: gap of unknown length
* 8986: contig of 1381 bp in length
* 10366: gap of unknown length
* 10466: contig of 1225 bp in length
* 11691: gap of unknown length
* 11791: contig of 1325 bp in length
* 13116: gap of unknown length
* 13216: contig of 1018 bp in length
* 14234: gap of unknown length
* 14335: contig of 1138 bp in length
* 15472: gap of unknown length
* 15573: contig of 1117 bp in length
* 16689: gap of unknown length
* 16690: contig of 1049 bp in length
* 17838: contig of 1049 bp in length

```
* 17839 17938: gap of unknown length
* 17939 18188: contig of 1250 bp in length
* 18189 18286: gap of unknown length
* 18289 20826: contig of 1538 bp in length
* 20827 21984: contig of 1058 bp in length
* 21985 22084: gap of unknown length
* 22085 23158: contig of 1074 bp in length
* 23159 23258: gap of unknown length
* 23259 24597: contig of 1339 bp in length
* 24598 24697: gap of unknown length
* 24698 25784: contig of 1087 bp in length
* 25785 25884: gap of unknown length
* 25885 26929: contig of 1045 bp in length
* 26930 28535: contig of 1506 bp in length
* 28536 28635: gap of unknown length
* 28636 30028: contig of 1393 bp in length
* 30029 31329: gap of unknown length
* 31330 31431: contig of 1203 bp in length
* 31432 32511: gap of unknown length
* 32512 32611: contig of 1080 bp in length
* 32611 33889: contig of 1278 bp in length
* 33890 33989: gap of unknown length
* 33990 35192: contig of 1203 bp in length
* 35193 35292: gap of unknown length
* 35293 36574: contig of 1282 bp in length
* 36575 36674: gap of unknown length
* 36675 37944: contig of 1270 bp in length
* 37945 38044: gap of unknown length
* 38045 39129: contig of 1085 bp in length
* 39130 39229: gap of unknown length
* 39230 40571: contig of 1342 bp in length
* 40572 40671: gap of unknown length
* 40672 42172: contig of 1501 bp in length
* 42173 42272: gap of unknown length
* 42273 44167: contig of 1895 bp in length
* 44168 44267: gap of unknown length
* 44268 45726: contig of 1459 bp in length
* 45727 45826: gap of unknown length
* 45827 47489: contig of 1663 bp in length
* 47490 47589: gap of unknown length
* 47590 48678: contig of 1289 bp in length
* 48679 48978: gap of unknown length
* 48979 50766: contig of 1788 bp in length
* 50767 50866: gap of unknown length
* 50867 54237: contig of 3371 bp in length
* 54238 54337: gap of unknown length
* 54338 56520: contig of 2183 bp in length
* 56521 56620: gap of unknown length
* 56621 58133: contig of 1513 bp in length
* 58134 59233: gap of unknown length
* 59234 59527: contig of 1294 bp in length
* 59528 59627: gap of unknown length
* 59628 63467: contig of 3840 bp in length
* 63468 63567: gap of unknown length
* 63568 65727: contig of 2160 bp in length
* 65728 65827: gap of unknown length
* 65828 68787: contig of 2960 bp in length
* 68788 72316: gap of unknown length
* 72317 72416: gap of unknown length
* 72417 76575: contig of 4159 bp in length
* 76576 76675: gap of unknown length
* 76676 82272: contig of 5597 bp in length
* 82273 86199: contig of 3827 bp in length
* 86200 86299: gap of unknown length
* 86300 91846: contig of 5547 bp in length
* 91847 91946: gap of unknown length
* 91947 96394: contig of 4448 bp in length
* 96395 96494: gap of unknown length

* 96495 102483: contig of 5989 bp in length
* 102484 102583: gap of unknown length
* 102584 110476: contig of 7893 bp in length
* 110477 110576: gap of unknown length
* 110577 119241: contig of 8665 bp in length
* 119242 119341: gap of unknown length
* 119342 130059: contig of 10718 bp in length
* 130060 130159: gap of unknown length
* 130160 142204: contig of 12045 bp in length
* 142205 142304: gap of unknown length
* 142305 157417: contig of 15113 bp in length
* 157418 157517: gap of unknown length
* 157518 185392: contig of 27875 bp in length
* 185393 185492: gap of unknown length
* 185493 212309: contig of 26817 bp in length
* 212310 212409: gap of unknown length
* 212410 249959: contig of 37550 bp in length
* 249960 250059: gap of unknown length
* 250060 279589: contig of 29530 bp in length.

FEATURES             Location/Qualifiers
     source            1..279589
                     /organism="Mus musculus"
                     /db_xref="taxon:10090"
                     /clone_lib="RP23-359N18"
                     /clone_lib="RPCI mouse BAC library 23"
BASE COUNT      69254 a 68603 c 67231 g 68544 t 5957 others
ORIGIN

alignment_scores:    Quality: 954.00      Length: 210
                    Ratio: 4.654         Gaps: 4
Percent Similarity: 97.619      Percent Identity: 95.714

alignment_block:
US-09-466-935-4 x AC073768/rev ..

Align seg 1/1 to reverse of: AC073768 from: 1 to: 279589

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
30830 ATGTTGATGTTATTTCACCGTCGCCATGGTGACATTGTGGCGCTTAT 30781
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
30780 GAGCCCGGTCGCCATTCTTTTGTGCTCTCAGACCGCTGTCAGTCGTT 30731
34 eArArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
30730 CCGGTAAAGAAAGCATGATGGCGTGTGGGCATTACCTGCGGCGTAATG 30681
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMe 67
30680 GTTGGGCTGGGATTGGCTGCTGGCTGCATTGATTATCGAANAAT 30631
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyrLeuCysT 84
30630 GCGCTGGCTGCATACGCTGATTATGGTGGCGGTGGCTGTATCTCTGCT 30581
84 tPmMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAla.Valse 100
30580 GGATGGGTTACCAGATGCTACGTGGTGTCACTGAAAGAGGCGCGTTTC 30531
100 tAlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuL 117
30530 TGCACCTCGCCACAGCTGAGCTGGCGAAGAGTGGCGCAGTTTCCTGA 30481
117 ysgGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGly 133
30480 AAGGTTTACTGACCATCGCTAAATCCCGAAGCGATTATCTACTTTGCG 30431
134 SerValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTr 150
|||||
```

```
30430 TCGGTGTTCTCATTTGTTGCGTGATACGTTGGCACTACCGCGCTG 30381
150 pGlyIlePheAlaLeuIle.IleValGluThrLeuAlaTrpPhehrVal 166
|||||
30380 GGGCATTTTTCGCTGATCATTTGCGAAACGCTGGCGTGGTTTACCGTC 30331
167 Val.AlaserLeu.PheAlaLeuProGlnMetArgArgGlyTyrGlnArg 182
|||||
30330 GTTCGGCAGCTGCTGCTGCGCTGCCCAATGCGCGTGGTTATCA.CGT 30282
183 LeuAlaIlysrIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheG1 199
|||||
30281 CTGCGGAAGTGGATTGATGCTTTTTCGCGGCGCTATTTCGCGGATTGG 30232
199 YlleHisIleIleSerArg 206
|||||
30231 CATTCATTGATTATTCGCGG 30210
seq_name: gb_htg7:AC020970

seq_documentation_block:
LOCUS AC020970 212936 bp DNA HTG 10-FEB-2000
DEFINITION Mus musculus clone Rp23-252M21, WORKING DRAFT SEQUENCE, 144
unordered pieces.
ACCESSION AC020970
VERSION AC020970.1 GI:6691260
KEYWORDS HTG; HTGS-PHASE1; HTGS-DRAFT.
SOURCE house mouse.
ORGANISM Mus musculus
Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
1 (bases 1 to 212936)
DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
AUTHORS DOE Joint Genome Institute.
REFERENCE 2 (bases 1 to 212936)
JOURNAL Direct Submission
TITLE Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----Summary Statistics
Estimated insert size: 212936; sum-of-contigs estimation
Estimated insert size: 157000; pulse field gel estimation
Quality coverage: 4.20x in Q20 bases; pulse field gel estimation
Quality coverage: 3.10x in Q20 bases; sum-of-contigs estimation
-----
* NOTE: This is a 'working draft' sequence. It currently
* consists of 144 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.
1
* 1159: contig of 1159 bp in length
* gap of unknown length
* 1160 2363: contig of 1204 bp in length
* gap of unknown length
* 2364 3604: contig of 1241 bp in length
* gap of unknown length
* 3605 5188: contig of 1584 bp in length
* gap of unknown length
* 5189 6372: contig of 1184 bp in length
* gap of unknown length
* 6373 7737: contig of 1365 bp in length
* gap of unknown length
* 7738 9783: contig of 2046 bp in length
* gap of unknown length
* 9784 11046: contig of 1263 bp in length
```

* 58687 60407: contig of 1721 bp in length
* gap of unknown length
* 60408 61632: contig of 1225 bp in length
* gap of unknown length
* 61633 62880: contig of 1248 bp in length
* gap of unknown length
* 62881 64220: contig of 1340 bp in length
* gap of unknown length
* 64221 65420: contig of 1200 bp in length
* gap of unknown length
* 65421 66605: contig of 1185 bp in length
* gap of unknown length
* 66606 68682: contig of 2077 bp in length
* gap of unknown length
* 68683 70712: contig of 2030 bp in length
* gap of unknown length
* 70713 71895: contig of 1183 bp in length
* gap of unknown length
* 71896 73462: contig of 1567 bp in length
* gap of unknown length
* 73463 74806: contig of 1344 bp in length
* gap of unknown length
* 74807 76205: contig of 1399 bp in length
* gap of unknown length
* 76206 77580: contig of 1375 bp in length
* gap of unknown length
* 77581 79081: contig of 1501 bp in length
* gap of unknown length
* 79082 80314: contig of 1233 bp in length
* gap of unknown length
* 80315 81899: contig of 1585 bp in length
* gap of unknown length
* 81900 83140: contig of 1241 bp in length
* gap of unknown length
* 83141 84665: contig of 1525 bp in length
* gap of unknown length
* 84666 85888: contig of 1223 bp in length
* gap of unknown length
* 85889 87155: contig of 1267 bp in length
* gap of unknown length
* 87156 88435: contig of 1280 bp in length
* gap of unknown length
* 88436 89955: contig of 1520 bp in length
* gap of unknown length
* 89956 91277: contig of 1322 bp in length
* gap of unknown length
* 91278 92513: contig of 1236 bp in length
* gap of unknown length
* 92514 93917: contig of 1404 bp in length
* gap of unknown length
* 93918 95448: contig of 1531 bp in length
* gap of unknown length
* 95449 96800: contig of 1352 bp in length
* gap of unknown length
* 96801 97978: contig of 1178 bp in length
* gap of unknown length
* 97979 99179: contig of 1201 bp in length
* gap of unknown length
* 99180 100437: contig of 1258 bp in length
* gap of unknown length
* 100438 101827: contig of 1390 bp in length
* gap of unknown length
* 101828 103117: contig of 1290 bp in length
* gap of unknown length
* 103118 104550: contig of 1433 bp in length
* gap of unknown length
* 104551 105854: contig of 1304 bp in length
* gap of unknown length
* 105855 107523: contig of 1669 bp in length
* gap of unknown length
* 107524 109208: contig of 1685 bp in length
* gap of unknown length
* 109209 110430: contig of 1222 bp in length

* 110431 111662: contig of 1232 bp in length
* gap of unknown length
alignment_scores:
Quality: 747.00 Length: 206
Ratio: 3.698 Gaps: 0
Percent Similarity: 98.058 Percent Identity: 96.602
alignment_block:
US-09-466-935-4 x AC020970 ..
Align seg 1/1 to: AC020970 from: 1 to: 212936
1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
174261 ATGTTGATGTTATTTCTCACCGTCGCATGGTGCACATTTGGCGCTTAT 174310
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArg 34
174311 GAGCCCGGTCGCCGATTTCCTT.TTTGCTCTCAGACCGCTGTCAGTCGTT 174359
34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysClyValMet 50
174360 CCCGTAAAGAAGCGATGAT.GGCGTCTGGCATTACCTGCGCGTAATG 174408
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMe 67
174409 GTTTGGGCTGGATTGGCT.CTTGGCCTGCATTTGATTATCGAAAAAT 174457
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuYrLeuCyst 84
174458 GGCCTGCTGCATACGCTGAT.ATGCTGGGCGGTGGCTGTATCTCTGCT 174506
84 rpMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
174507 GGATGGGTTACAGATGCTAGT.GGTGCACTGAAAAAGAGCGGTTCT 174555
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLy 117
174556 GCACCTGCGCACAGGTGCGA.CTGGCGAAAAAGTGGCGCAGTTCTCTGAA 174604
117 sGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleIleIleIle 134
174605 AGGTTTACTGACCAATCTCGT.AATCCGAAAGCGATTATCTACTTGGCT 174653
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrp 150
174654 CGGTGTTCTCATTTGTCGT.GATAACGTTGGCAGTACCGCGCGCTGG 174702
151 GlyIlePheAlaLeuIleIleValGluThrLeuAlaTrpPheThrValVa 167
174703 GGCATTTTGGCTGATCAT.GTCGAAACGCTGGCGTGTACCGTCGT 174751
167 lAlaSerLeuPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuA 184
174752 TGCAGCCTGTTGCCCTGGG.CAAATGCGCGCTGTTATCAACGCTCGG 174800
184 laLysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
174801 CGAAGTGGATTGATGTTTTC.CGGGCGGTATTATTCGCGGATTTGGCATT 174849
201 HisLeuIleIleSerArg 206
174850 CATTTGATTATTCGCGG 174867
seq_name: gb_htg7:AC020885
seq_documentation_block:
LOCUS AC020885 268294 bp DNA HTG 16-FEB-2000
DEFINITION Mus musculus clone RP23-464I1, LOW-PASS SEQUENCE SAMPLING.
ACCESSION AC020885
VERSION AC020885.2 GI:6980212
KEYWORDS HTG; HTGS_PHASE0.


```

* 44512 45780: gap of unknown length
* 45781 47202: contig of 1269 bp in length
* 47203 48647: contig of 1422 bp in length
* 48648 49652: contig of 1445 bp in length
* 49653 50485: contig of 1005 bp in length
* 50486 51527: contig of 833 bp in length
* 51528 53119: contig of 1042 bp in length
* 53120 53623: contig of 1592 bp in length
* 53624 54669: contig of 504 bp in length
* 54670 55309: contig of 1046 bp in length
* 55310 56382: contig of 504 bp in length
* 56383 56626: contig of 1060 bp in length
* 56626 57345: contig of 640 bp in length
* 57346 57505: contig of 1073 bp in length
* 57506 58908: contig of 243 bp in length
* 58909 59968: contig of 720 bp in length
* 59969 61377: contig of 160 bp in length
* 61378 62292: contig of 1403 bp in length
* 62292 63319: contig of 1060 bp in length
* 63320 63818: contig of 1409 bp in length
* 63819 64823: contig of 914 bp in length
* 64824 65269: contig of 1028 bp in length
* 65270 65904: contig of 499 bp in length
* 65905 66815: contig of 1005 bp in length
* 66815 6815: contig of 446 bp in length
* 6815 6915: contig of 635 bp in length
* 6915 7015: contig of 911 bp in length
* 7015 7115: gap of unknown length

alignment_scores:
  Quality: 717.00      Length: 140
  Ratio: 5.158        Gaps: 0
  Percent Similarity: 99.286      Percent Identity: 98.571

alignment_block:
  US-09-466-935-4 x AC020885/rev ..

Align seg 1/1 to reverse of: AC020885 from: 1 to: 268294

67 MetAlaTrpLeuHisThrLeuIleMetValGlyGlyLeuTyrLeuCy 83
::: |||||
11743 CTGCTTGGCTGCATACGCTGATATAGTGGCGGTGGCCCTGATCTCTG 11694

83 sTrpMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValS 100
|||||
11693 CTGGATGGGTTACCAAGATGCTACGTGGTGCACGTGAATAAGAGCGGTTT 11644

100 eRAlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeu 116
|||||
11643 CTGCACCTGGCCACAGGTCGAGCTGGCGAAAGTGGCGAGCTTCTCTG 11594

117 LysGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIlelleTyrPheGl 133
```

```

|||||
11593 AAAGGTTTACTGACCAATCTCGTAACTCCGAAGCGATTATCTACTTTGG 11544

133 ySerValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgT 150
|||||
11543 CTGGTGTCTCTCATTTGTTGCGTGATAGCTGGCAGCTACCGCGGCT 11494

150 rPGLyIlePheAlaLeuIleValGluThrLeuAlaTrpPheThrVal 166
|||||
11493 GGGCATTTTGGCGTGAATGTCGAAACGCTGGCGTGGTTTACCGTC 11444

167 ValAlaSerLeuPheAlaLeuProGlnMetArgGlyTyrGlnArgLe 183
|||||
11443 GTTGCAGGCTGTTTGCCTGCCGCAATGCGCGTGGTATCAACGCT 11394

183 uAlaLysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyI 200
|||||
11393 GCGCAAGTGGATGATGTTTTCGCGGGCGCTTATTTGCCGGATTGGCA 11344

200 leHisLeuIlelleSerArg 206
|||||
11343 TTCAATTTGATTATTTGCGGG 11324

seq_name: gb_htgl7:AC079167

seq_documentation_block:
  LOCUS AC079167 290452 bp DNA HTG 30-AUG-2000
  DEFINITION Mus musculus chromosome 6 clone RP23-226D23 strain C57BL6/J, ***
  SEQUENCING IN PROGRESS ***, 196 unordered pieces.
  ACCESSION AC079167
  VERSION AC079167.2 GI:9945047
  KEYWORDS HTG; HTGS_PHASE1.
  SOURCE house mouse.
  ORGANISM Mus musculus
    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
  REFERENCE 1 (bases 1 to 290452)
  AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
    Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
    Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE High Throughput Mouse Sequencing
  JOURNAL Unpublished
  REFERENCE 2 (bases 1 to 290452)
  AUTHORS Grills,G., Han,J., Montgomery,K.T., Lee,E., Long,J., Pomerantz,R.,
    Ioshikhes,I.P., Shim,C., Decker,J., Thomas,E., Perera,A.,
    Gordon,M., Goltz,J.S. and Kucherlapati,R.
  TITLE Direct Submission
  JOURNAL Submitted (23-AUG-2000) Department of Molecular Genetics, Albert
    Einstein College of Medicine Genome Center, 1300 Morris Park Ave.,
    Bronx, NY 10461, USA
  COMMENT On Aug 30, 2000 this sequence version replaced gi:9885883.
    -----Genome Center:
    Albert Einstein College of Medicine Center
    Code: AECOM
    Web site:
    http://sequence.aecom.yu.edu/cgi-
    bin/ws.exe/mouseDB/mouseSeq/mouseseqtable.hts
    Contact: jhan@sequence.aecom.yu.edu
    -----Summary Statistics
    Center project name: AEN
    Sequencing vector: pUC18; L08752
    Chemistry: Dye-terminator Big Dye; 100%
    *Consensus quality: 196511 at least Q20
    *Consensus quality: 172308 at least Q30
    *Consensus quality: 138790 at least Q40
    Estimated insert size: agarose-FP - N/A
    **Estimated insert size: 286552 - sum-of-contents
    Quality coverage: agarose-FP - N/A
    Quality coverage: 3.3x sum-of-contents - N/A
    -----
    * NOTE: This is a 'working draft' sequence. It currently
    * consists of 196 contigs. The true order of the pieces
    * is not known and their order in this sequence record is
```

* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence.
* as soon as it is available and the accession number will
* be preserved.

* 852:	contig of 852 bp in length		
* 872:	gap of unknown length		
* 873:	contig of 881 bp in length	30181	31014: contig of 834 bp in length
* 1754:	gap of unknown length	31015	31034: gap of unknown length
* 1774:	contig of 885 bp in length	31035	31880: contig of 846 bp in length
* 2679:	gap of unknown length	31881	31900: gap of unknown length
* 2679:	gap of unknown length	31901	33160: contig of 1260 bp in length
* 3514:	contig of 836 bp in length	33161	33180: gap of unknown length
* 3534:	gap of unknown length	33181	34033: contig of 853 bp in length
* 3535:	contig of 915 bp in length	34034	34053: gap of unknown length
* 4450:	gap of unknown length	34054	34972: contig of 919 bp in length
* 5384:	contig of 915 bp in length	34973	34992: gap of unknown length
* 5404:	gap of unknown length	34993	35875: contig of 883 bp in length
* 6273:	contig of 869 bp in length	35876	35895: gap of unknown length
* 6274:	gap of unknown length	35896	36731: contig of 836 bp in length
* 6294:	contig of 802 bp in length	36732	36751: gap of unknown length
* 7095:	contig of 802 bp in length	36752	37639: contig of 888 bp in length
* 7115:	gap of unknown length	37640	37659: gap of unknown length
* 7881:	contig of 866 bp in length	37660	38499: contig of 840 bp in length
* 8001:	gap of unknown length	38500	38519: gap of unknown length
* 8002:	contig of 728 bp in length	39349	39348: contig of 829 bp in length
* 8749:	gap of unknown length	39369	39368: gap of unknown length
* 8750:	contig of 1126 bp in length	40378	40377: contig of 1009 bp in length
* 8750:	contig of 1126 bp in length	40378	40397: gap of unknown length
* 8750:	contig of 1126 bp in length	40398	41469: contig of 1072 bp in length
* 8750:	contig of 1126 bp in length	41470	41489: gap of unknown length
* 8750:	contig of 1126 bp in length	41490	41890: contig of 401 bp in length
* 8750:	contig of 1126 bp in length	41891	41910: gap of unknown length
* 8750:	contig of 1126 bp in length	41911	42795: contig of 885 bp in length
* 8750:	contig of 1126 bp in length	42796	42815: gap of unknown length
* 8750:	contig of 1126 bp in length	42816	43744: contig of 929 bp in length
* 8750:	contig of 1126 bp in length	43745	43764: gap of unknown length
* 8750:	contig of 1126 bp in length	43765	44638: contig of 864 bp in length
* 8750:	contig of 1126 bp in length	44639	44648: gap of unknown length
* 8750:	contig of 1126 bp in length	44649	45958: contig of 1310 bp in length
* 8750:	contig of 1126 bp in length	45959	45978: gap of unknown length
* 8750:	contig of 1126 bp in length	45979	46827: contig of 849 bp in length
* 8750:	contig of 1126 bp in length	46828	46847: gap of unknown length
* 8750:	contig of 1126 bp in length	46848	47807: contig of 960 bp in length
* 8750:	contig of 1126 bp in length	47808	47827: gap of unknown length
* 8750:	contig of 1126 bp in length	47828	48376: contig of 549 bp in length
* 8750:	contig of 1126 bp in length	48377	48396: gap of unknown length
* 8750:	contig of 1126 bp in length	48397	49297: contig of 901 bp in length
* 8750:	contig of 1126 bp in length	49298	49317: gap of unknown length
* 8750:	contig of 1126 bp in length	49318	50182: contig of 865 bp in length
* 8750:	contig of 1126 bp in length	50183	50202: gap of unknown length
* 8750:	contig of 1126 bp in length	50203	51054: contig of 852 bp in length
* 8750:	contig of 1126 bp in length	51055	51074: gap of unknown length
* 8750:	contig of 1126 bp in length	51075	51373: contig of 299 bp in length
* 8750:	contig of 1126 bp in length	51374	51393: gap of unknown length
* 8750:	contig of 1126 bp in length	51394	52053: contig of 660 bp in length
* 8750:	contig of 1126 bp in length	52054	52073: gap of unknown length
* 8750:	contig of 1126 bp in length	52074	52944: contig of 871 bp in length
* 8750:	contig of 1126 bp in length	52945	52964: gap of unknown length
* 8750:	contig of 1126 bp in length	52965	53539: contig of 575 bp in length
* 8750:	contig of 1126 bp in length	53540	53559: gap of unknown length
* 8750:	contig of 1126 bp in length	53560	54374: contig of 815 bp in length
* 8750:	contig of 1126 bp in length	54375	54394: gap of unknown length
* 8750:	contig of 1126 bp in length	54395	55253: contig of 859 bp in length
* 8750:	contig of 1126 bp in length	55254	55273: gap of unknown length
* 8750:	contig of 1126 bp in length	55274	56527: contig of 1254 bp in length
* 8750:	contig of 1126 bp in length	56528	56547: gap of unknown length
* 8750:	contig of 1126 bp in length	56548	57376: contig of 829 bp in length
* 8750:	contig of 1126 bp in length	57377	57396: gap of unknown length
* 8750:	contig of 1126 bp in length	57397	57473: contig of 77 bp in length
* 8750:	contig of 1126 bp in length	57474	57493: gap of unknown length
* 8750:	contig of 1126 bp in length	57494	58337: contig of 844 bp in length
* 8750:	contig of 1126 bp in length	58338	58357: gap of unknown length
* 8750:	contig of 1126 bp in length	58358	58444: contig of 87 bp in length
* 8750:	contig of 1126 bp in length	58445	58464: gap of unknown length
* 8750:	contig of 1126 bp in length	58465	59288: contig of 824 bp in length
* 8750:	contig of 1126 bp in length	59289	59308: gap of unknown length
* 8750:	contig of 1126 bp in length	59290	60499: contig of 1191 bp in length
* 8750:	contig of 1126 bp in length	60500	61859: contig of 1340 bp in length
* 8750:	contig of 1126 bp in length	60501	
* 8750:	contig of 1126 bp in length	60502	
* 8750:	contig of 1126 bp in length	60503	
* 8750:	contig of 1126 bp in length	60504	
* 8750:	contig of 1126 bp in length	60505	
* 8750:	contig of 1126 bp in length	60506	
* 8750:	contig of 1126 bp in length	60507	
* 8750:	contig of 1126 bp in length	60508	
* 8750:	contig of 1126 bp in length	60509	
* 8750:	contig of 1126 bp in length	60510	
* 8750:	contig of 1126 bp in length	60511	
* 8750:	contig of 1126 bp in length	60512	
* 8750:	contig of 1126 bp in length	60513	
* 8750:	contig of 1126 bp in length	60514	
* 8750:	contig of 1126 bp in length	60515	
* 8750:	contig of 1126 bp in length	60516	
* 8750:	contig of 1126 bp in length	60517	
* 8750:	contig of 1126 bp in length	60518	
* 8750:	contig of 1126 bp in length	60519	
* 8750:	contig of 1126 bp in length	60520	

* 61860 61879: gap of unknown length
* 61880 62752: contig of 873 bp in length
* 62753 62772: gap of unknown length
* 62773 63623: contig of 851 bp in length
* 63624 63643: gap of unknown length
* 63644 64694: contig of 1051 bp in length
* 64695 64714: gap of unknown length

alignment_scores:
Quality: 559.00 Length: 116
Ratio: 4.861 Gaps: 0
Percent Similarity: 99.138 Percent Identity: 99.138

alignment_block:

US-09-466-935-4 x AC079167

Align seg 1/1 to: AC079167 from: 1 to: 290452

91 ArgGlyAlaLeuLysGlyLeuAlaValSerAlaProAlaProGlnVal1 107
|||||
104050 CGTGGTCACTGAAAAAGAGCG.GTTTCTGCACCTGCCACAGTCCA 104098

107 uLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuLeuThrAsnLeuA 124
|||||

104099 GCTGGCGAAAAGTGGCGCAGTTCTCGAAAGTTTACTGACCAATCTCG 104148
|||||

124 laAsnProLysAlaIleIleTyPheGlySerValPheSerLeuPheVal 140
|||||

104149 CTAATCGGAAAGCGATATCTATTGGCTCGGTGTTCTCATTTGTTGTC 104198
|||||

141 GlyAspAsnValGlyThrAlaArgTrpGlyIlePheAlaLeuIle1 157
|||||

104199 GGTGATACCGTTGGCACTACCGCGCGCTGGGCGATTTTGGCGTATCAT 104248
|||||

157 eValGluThrLeuAlaTrpPheThrValValAlaSerLeuPheAlaLeuP 174
|||||

104249 TGTGGAACGCTGGCGGTGTTTACCGCTGTTGCCAGCCTGTTTGGCCCTGC 104298
|||||

174 roGlnMetArgArgGlyTyGlnArgLeuAlaLysTrpIleAspGlyPhe 190
|||||

104299 CGCAAAATGCGCGGTGTTATCAACGCTCTGGCGAAGTGGATTGATGTTT 104348
|||||

191 AlaGlyAlaLeuPheAlaGlyPheGlyIleHisLeuIleIleSerArg 206
|||||

104349 GCCGGGCGGTATTTCGGGATTGGCAATTCATTGATTATTTCGCGG 104396
|||||

seq_name: gb_ba2:ECORECQ

seq_documentation_block:

LOCUS ECORECQ 2695 bp DNA BCT 26-APR-1993
DEFINITION E.coli recQ gene complete cds, and pldA gene, 3' end.
ACCESSION M30198
VERSION M30198.1 GI:147557
KEYWORDS detergent-resistant phospholipase A; phospholipase; pldA gene; recQ gene.

SOURCE Escherichia coli (strain K-12) DNA.

ORGANISM Escherichia coli

Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae; Escherichia.

REFERENCE 1 (bases 1 to 2695)

AUTHORS Irino,N., Nakayama,K. and Nakayama,H.

TITLE The recQ gene of Escherichia coli K12: Primary structure and

evidence for SOS regulation

JOURNAL Mol. Gen. Genet. 205, 298-304 (1986)

MEDLINE 87115164

FEATURES Location/Qualifiers

1..2695

/organism="Escherichia coli"

/strain="K-12"

/db_xref="taxon:562"

1..352

/gene="pldA"

<1..352

gene

CDS

/gene="pldA"
/codon_start=2
/transl_table=11
/product="detergent-resistant phospholipase A"
/protein_id="AAA24516.1"
/db_xref="GI:147558"
/translation="SWNRLYTRLMAENGWLEVEKPKWYVVGNTDDNPDIKMGYYQL
KIGYHLGDAVLSAKQYNNWTGYGAEGLGLSYPTTKHVRLYTQVYSGYGESLIDYFN
QTRVGVGVMNLDF"
479..2311
/gene="recQ"
479..2311
/gene="recQ"
/codon_start=1
/transl_table=11
/protein_id="AAA24517.1"
/db_xref="GI:147559"
/translation="MNVAAEVLNLESGAKVLOETGYQOFRPQREIIDTVLSGRD
CLVMPGTGGKSLCYOIPALLLGLTVVVSPLISLMDQDQDQQAANGVAAACLNSTQT
REQLVMTGCTGTQTRLLYIAPERLMDNLEHLAHWNVLLAVDEAHCISQWGHDF
RPEYALGQLRQRPFTLPFALTATADTTQDILVRLGLNDPLIQISSFDRENIRYM
LMEKFKPLDMRYVQEQRGSGIYICNSRAKVEDTAAALQSKGISAAYHAGLENNY
RADVQKFORDDLIWVAFVGMGINKPNRVFVVFDDIPRNIESYQETGRAGRDGL
PAEAMLFYDPADMAWLRCLBEKPGQLODIERHKLNAMGAFEAQTCRRLVLLNVEG
EGRECEGNCDCILDPPKOYDSDAOIALSTIGRVNOREGMSVYVVEVIRGANNQIR
DYGHDKLVYGMGRDKSHEHWVSIVRILHLGLVTONIAQHSALQITEARPVLAESS
LQLAVPRIVALAPKAMQKSGNDRKLFKLRKLKRSIADESNNPVPYVFNDAITLIE
MAEOMPITASEMLSVNGVMKRLERFGFPFPMALIPAHVDGDDDEE"
BASE COUNT 618 a 680 c 764 g 633 t
ORIGIN

alignment_scores:
Quality: 486.00 Length: 107
Ratio: 4.629 Gaps: 0
Percent Similarity: 98.131 Percent Identity: 98.131

alignment_block:

US-09-466-935-4 x ECORECQ

Align seg 1/1 to: ECORECQ from: 1 to: 2695

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
|||||

2375 ATGTTGATGTTATTCTCACCGTCGCCATGGTGACATTTGGCGCTTAT 2424
|||||

17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArg 34
|||||

2425 GAGCCCCGTCGCCGATTTCTTTTGTCTCTCAGACCGCTGTCTCAGTCGT 2474
|||||

34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
|||||

2475 CCGGTAAAGACCGATGATGGCGCTGTGGGCATTAC.TGCGGCGTAATG 2523
|||||

51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMe 67
|||||

2524 GTTTGGGCTGGGATTGCGCTGTTGGCTGCATTGATTATCGAAAAAAT 2573
|||||

67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyTrpLeuCys 84
|||||

2574 GCCCTGGCTCGATACGCTG.ATTATGTTGGCGGTGGCGCTGTATCTCTGCT 2622
|||||

84 rpMetGlyTyTrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
|||||

2623 GGATGGGTACCATGCTACGTGCTGCTACTGAAAGAGGCGGTTTCT 2672
|||||

101 AlaProAlaProGlnValGlu 107
|||||

2673 GCACCTGCCACAGTTCGAG 2693

seq_name: gb_htg7:AC020833

seq_documentation_block:

LOCUS AC020833 62274 bp DNA HTG 16-FEB-2000
DEFINITION Mus musculus clone RG-WBAC_40H2, LOW-PASS SEQUENCE SAMPLING.
AC020833
AC020833.1 GI:6686475
VERSION HTG; HTGS_PHASE0.
KEYWORDS house mouse.
SOURCE Mus musculus
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Mouse
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 62274)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (10-JAN-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT
* NOTE: This record contains 69 individual
* sequencing reads that have not been assembled into
* contigs. Runs of N are used to separate the reads
* and the order in which they appear is completely
* arbitrary. Low-pass sequence sampling is useful for
* identifying clones that may be gene-rich and allows
* overlap relationships among clones to be deduced.
* However, it should not be assumed that this clone
* will be sequenced to completion. In the event that
* the record is updated, the accession number will
* be preserved.
1
296: contig of 296 bp in length
297 gap of unknown length
1049: contig of 753 bp in length
1050 gap of unknown length
1812: contig of 763 bp in length
1813 gap of unknown length
2420: contig of 608 bp in length
2421 gap of unknown length
2953: contig of 533 bp in length
2954 gap of unknown length
3645: contig of 692 bp in length
3646 gap of unknown length
4766: contig of 1121 bp in length
4767 gap of unknown length
5296: contig of 530 bp in length
5297 gap of unknown length
5584: contig of 288 bp in length
5585 gap of unknown length
6454: contig of 870 bp in length
6455 gap of unknown length
6853: contig of 399 bp in length
6854 gap of unknown length
7566: contig of 713 bp in length
7567 gap of unknown length
8253: contig of 687 bp in length
8254 gap of unknown length
9058: contig of 805 bp in length
9059 gap of unknown length
9678: contig of 620 bp in length
9679 gap of unknown length
10272: contig of 594 bp in length
10273 gap of unknown length
10890: contig of 618 bp in length
10891 gap of unknown length
11695: contig of 805 bp in length
11696 gap of unknown length
12677: contig of 982 bp in length
12678 gap of unknown length
13130: contig of 453 bp in length
13131 gap of unknown length
13878: contig of 748 bp in length
14026: contig of 148 bp in length
gap of unknown length
14027 contig of 843 bp in length
14869: gap of unknown length
15738: contig of 869 bp in length
15739 gap of unknown length
16427 contig of 688 bp in length
17161: gap of unknown length
17162 contig of 735 bp in length
17162 gap of unknown length
17938: contig of 777 bp in length
17939 gap of unknown length
18715: contig of 777 bp in length
18716 gap of unknown length
19493: contig of 778 bp in length
19494 gap of unknown length
20212: contig of 719 bp in length
20213 gap of unknown length
20780: contig of 568 bp in length
20781 gap of unknown length
21499: contig of 719 bp in length
21500 gap of unknown length
21571: contig of 71 bp in length
22378: contig of 808 bp in length
22379 gap of unknown length
22940: contig of 562 bp in length
22941 gap of unknown length
23943: contig of 1003 bp in length
23944 gap of unknown length
24650: contig of 707 bp in length
24651 gap of unknown length
25686: contig of 1036 bp in length
25687 gap of unknown length
25820: contig of 134 bp in length
25821 gap of unknown length
26268: contig of 448 bp in length
26269 gap of unknown length
27152: contig of 884 bp in length
27153 gap of unknown length
28057: contig of 905 bp in length
28058 gap of unknown length
29002: contig of 945 bp in length
29003 gap of unknown length
29991: contig of 989 bp in length
29992 gap of unknown length
30780: contig of 789 bp in length
30781 gap of unknown length
31514: contig of 734 bp in length
31515 gap of unknown length
32152: contig of 638 bp in length
32153 gap of unknown length
32933: contig of 781 bp in length
32934 gap of unknown length
33042: contig of 109 bp in length
33043 gap of unknown length
34310: contig of 1268 bp in length
34311 gap of unknown length
34992: contig of 682 bp in length
34993 gap of unknown length
35821: contig of 829 bp in length
35822 gap of unknown length
37013: contig of 1192 bp in length
37014 gap of unknown length
38147: contig of 1134 bp in length
38148 gap of unknown length
38852: contig of 705 bp in length
38853 gap of unknown length
39840: contig of 988 bp in length
39841 gap of unknown length
40874: contig of 1034 bp in length
40875 gap of unknown length
42387: contig of 1513 bp in length
42388 gap of unknown length
43258: contig of 871 bp in length


```

/protein_id="AAK02811.1"
/db_xref="GI:12721022"
/translation="MMNLNLIIVHFFGLITPGDPFFVSVSLAASNRNRNALCAVIGITL
GVFWALASILGLAIFNTVTVLQGLVMTLGGGYLAYLGYLMKQQNVVFPVSEQE
QNKQTSIKETTKGLLVNLGNARAVIYFASVMSLVNLVNTQWQIWSALLIIVLETF
YFATISVFSRQQAQFYSOYSRYIDHLSGVIFFLLGGVLYMSGIGENTPLISK"
complement(3944..4429)
/gene="pgpA"
/note="PM0728"
complement(3944..4429)
/gene="pgpA"
/codon_start=1
/transl_table=1
/product="PgpA"
/protein_id="AAK02812.1"
/db_xref="GI:12721023"
/translation="MLKTPLOQLKSNPHFLALGFAGLIRPAGTWCMSVMGLMGW
GLQYLVNLSGFFELTALCFALGCVLCQKTAEDMGVHDGAIYVDEIVGIFVVLSLPN
LSPFWCLTAFILREFFDLPYPIRYFDHKLKESFGIMLDDILAAIYAVIYFILRW
P"
complement(4439..5431)
/gene="Chil"
/note="PM0729"
complement(4439..5431)
/gene="thil"
/codon_start=1
/transl_table=1
/product="Thil"
/protein_id="AAK02813.1"
/db_xref="GI:12721024"
/translation="MSHGPDILQVFTASKRTARKDVILSIGDCAITELPQNRVL
ITDPNVENTHFLPSMRADLAQVATNLNLDLAAGAPAWVSLALITPOVDAQWLS
DESQFFVLDHYNDVLDIGDTKGPLSVITTAHGLIPKGLGCRASASVQWNTYIS
TIGDSAGLSLLSQQLPNSDHEFLIQRHLRPTVRLVGLLELAVSSLANAALISDG
FLADUDHLTQSCQAGVELDKPLPLSSLVNVEFGVQAEFTALSGEYELCTVPDS
NKAKLERALAYIGVDYTCVGQVRAISPQNKKRIKFORGNPDITIQSGFDHFK"
complement(5448..5882)
/gene="nusb"
/note="PM0730"
complement(5448..5882)
/gene="nusb"
/codon_start=1
/transl_table=1
/product="Nusb"
/protein_id="AAK02814.1"
/db_xref="GI:12721025"
/translation="MTQKHKKISPRRRARECAVQALYSWYVQNSPAAIELNPMAE
QDLKQVDYAFRRLFRTAENVDAVDNIMIPYLDREVSELDPIEKAILRLAVIELKFE
LDVPYKVYINAEIYAVKVGAEDESHKYVNGVLDKVPVLSRK"
complement(5888..6361)
/gene="ribH"
/note="PM0731"
complement(5888..6361)
/gene="ribH"
/codon_start=1
/transl_table=1
/product="RibH"
/protein_id="AAK02815.1"
/db_xref="GI:12721026"
/translation="MKVLEGAAAPNAKIAVVIARFNSFINESLLEGALDALKRLGQV
KENTILRVFGAELPLVRLAELSKYDGIYALGTVIRGGTAHFYVAGASGGLG
QVAMNADIPVAFGVLTENIQAERACTRAGNKAGAEALVALEVMVNLQAIDAA"
6575..7786
/gene="tyrP_1"
/note="PM0732"
6575..7786
/gene="tyrP_1"
/codon_start=1
/transl_table=1
/product="TyrP"
/protein_id="AAK02816.1"
/db_xref="GI:12721027"
/translation="MKNKTRGSALLVAGTTIGAGMLAMPLTSAEMGFYVTLILLFILM
```

```

GSLYSALLFVEVYQRAETKNAGIATLAEOYFGLPGRILATLSLVFMVYLAISAVTGG
GSLIAGVPPLEGHATPIISVFTVILGVFIVISTGAVDMLTRFLFMKILVAEFPVL
LMLPLVNGDMLAMPKDLFIISASPVFTTSGFHVIIIPSIYSNLDGNIIRRLRIAI
TGTAIPLVAYILWOLATHGVFPOAQFVQILNTDPTNLGLTIYQATESAIISHAMRL
PFTLAIITSPUGVSLDFDCLYDLKARVAKTRNRSVGLLTFPLPLIFALFYPGFVM
ALGAGOMFTFPGVLVLPVGNARARKRYPDLFPRVYMGGLTLFGLALILGILMNVPEL
IKAGYLPVIG"
complement(7842..8504)
/gene="PM0733"
complement(7842..8504)
/gene="PM0733"
/codon_start=1
/transl_table=1
/product="unknown"
/protein_id="AAK02817.1"
/db_xref="GI:12721028"
/translation="MSQHTKOVLSAFHPSAIRRYDPNKKVSEEDFHYILELGRLS
SSVSGEPHFVVDROLREOLKAVSWGMOITDDASVILFLAKKHARYSPVPHDI
LAKGLDETOMOKALAIYERFODITOVLDNERTLEDWACKOTYIALANMWTGAAMIG
IDSCPIEGFDYCAVNNQLLSGTGAEDPNEGVSVANTFGYRAKEVAKSRKPLTEITTW
LK"
8727..10106
/gene="htrA"
/note="PM0734"
8727..10106
/gene="htrA"
/codon_start=1
/transl_table=1
/product="htrA"
/protein_id="AAK02818.1"
/db_xref="GI:12721029"
/translation="MKRERLLTSIALGLSILSTPMISOASLPTVDQGVPSLAPML
EKVLPAVVISVEGKARNHSTFGQIPPEFRFGPDLFNDRAPNRFRGIGSVIIN
ADSKLRLVNNHVIDSADKIITVQLDQRELSAKVIGTDEUSVALIOIEKPNKLTALTI
ADSKLRLVQGLIGINTAIISPSGNAGTAFALPSNMANNVQOQILEFGEVRRGLGI
KGGELNADLAKAFDIEAQOGAFYSEVLPNSAEAKAGLKAQDVIVAMNGOKISSFAEMR
AKIATSGAKREIALTIRDGKTHQTKVTQLSDDTQADASHLLPALAGAEMSNHDDKG
VKVLITSVTPTSLAEQRLKKGDIITGVNRQKVENLQLRKILDAKPSAIALNIVRG
NTNFYLLIQ"
BASE COUNT 3070 a 2201 c 1937 g 2976 t
ORIGIN

alignment_scores:
    Quality: 370.50      Length: 205
    Ratio: 2.406        Gaps: 2
    Percent Similarity: 75.122   Percent Identity: 37.073

alignment_block:
US-09-466-935-4 x AE006110/rev ..

Align seg 1/1 to reverse of: AE006110 from: 1 to: 10184

4 LeuphLeuthrValAlaMetValHisLeValAlaLeuMetSerProgl 20
      ::::::::::::::::::::: ::::::::::::::: :::::::::::::::
3938 GTGATGTTAAATCTTATTATTGTACATTTTTTCGGTTTAATCACGCCAGG 3899

20 yProAspPhePheValSerGlnThrAlaValSerArgSerArgLysG 37
      ::::::::::::::::::::: ::::::::::::::: :::::::::::::::
3888 ACCGACATTTTTTATGTCAGTCGATTAGTCGCCAGTAATTCACGTCGTA 3839

37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
      ::::::::::::::::::::: ::::::::::::::: :::::::::::::::
3838 ATGCACITTTGTCGCGTCGATTGTTACCTTAGGGGTGTTGTTTGGSCA 3789

54 GlyIleAlaLeuLeuGlyLeuHisLeuIleLeuGluLysMetAlaTrpLe 70
      ::::::::::::::::::::: ::::::::::::::: :::::::::::::::
3788 TTAGCCTCTATTATTAGGTTTAGCCATATTATTAAATACGGTTTCCTGTTT 3739

70 whisThrLeuIleMetValGlyGlyLeuTyrLeuCysTrpMetGlyT 87
      ::::::::::::::::::::: ::::::::::::::: :::::::::::::::
3738 ACAGGGATTAGTCATGACTTTAGGCGGTGGGTATTTAGCCTACTTGGGTT 3689
```



```
gene      5508..5807
CDS       /gene="BH2884"
          5508..5807
          /gene="BH2884"
          /note="BH2884
          unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="BAB06603.1"
          /db_xref="GI:10175506"
          /translation="MSQNERDFFDLMFGRPPSTDAEENPOESTSGSTVEETKTE
          GETKQEETNAAPALPLTFEQMEHIFRLAQSLGPAKLSLSPYVKTIQQLISQKKD"
          complement(5903..6157)
          /gene="BH2885"
          complement(5903..6157)
          /gene="BH2885"
          /note="BH2885
          unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="BAB06604.1"
          /db_xref="GI:10175507"
          /translation="MNVTGPPYRQYIGTGERFYPAVLPLFAGLAVGPLLFGGKP
          IGGYGPAYGPGSYGNFGNPGNPGMYPGPSPQPPFGYK"
          complement(6226..6555)
          /gene="BH2886"
          complement(6226..6555)
          /gene="BH2886"
          /note="BH2886
          unknown"
          /codon_start=1
          /transl_table=11
          /protein_id="BAB06605.1"
          /db_xref="GI:10175508"

alignment_scores:
  Quality: 292.00      Length: 207
  Ratio: 2.147        Gaps: 2
  Percent Similarity: 65.700      Percent Identity: 32.367

alignment_block:
US-09-466-935-4 x AP001517 ..

Align seg 1/1 to: AP001517 from: 1 to: 294250

4 LeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerProG1 20
  ::::: ||| ::::: ::::: ||| ::::: ||| ::::: ||| ::::: |||
52063 ATGTCGTGAGGTATTTCATCGTCGGCTGTTAGCAGGAATGCTCCTGG 52112

20 yProaspPhePheValSerGlnThrAlaValSerArgSerArgLysG 37
  ::::: ||| ::::: ::::: ||| ::::: ||| ::::: ||| ::::: |||
52113 CCCAGACTCTTTATCGTCATGAAACAGCGCTAGGTTGGAGCGAGAG 52162

37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
  ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52163 TTGGTATCCTAACTTCTTTAGGAGTGTCTGCGCTAATCGTTCATATT 52212

54 GlyIleAlaLeuGlyLeuHisIleIleGluLysMetAlaTrpLe 70
  ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52213 ACATATACAGTACGATTGCTTTTAAATTGAACATATCCTGCTCT 52262

70 uHisThrLeuLeuMetValGlyGlyLeuTyrlLeuCysTrpMetGlyT 87
  ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52263 TTTTTTTACGATTCAACTATTAGGAGCTGCTTATTATTGTTAGGTT 52312

87 yrGlnMetLeuArgGlyAla.....LeuLysLysGluAlaValSerAla 101
  ::::: ::::: ||| ::::: ||| ::::: ||| ::::: |||
52313 TTCATGCCATACGCTCATCTCCCTTAAAGGAAGAGCTGAAATCGAA 52362

102 ProAlaProGlnValGluLeuAlaLysSerGlyArgSer..... 114
  ::::: ::::: ||| ::::: ||| ::::: |||
52363 GAAACACACCAATCCATCAACAAAGATAGTAATCATCCATCAAGG 52412
```

```
115 PheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleIer 131
  ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52413 TTTCAAAGAAGGGTTTATTACTAAGCTTTTAAATCCTAAAGCTGCTCTCT 52462

131 yrPheGlySerValPheSerLeuPheValGlyAspAsnValGlyThrThr 147
  ::||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52463 TTTTTTTAAAGTATTTTTCACAATTTATAACCCGCAACGGCTGATGG 52512

148 AlaArgTrpGlyIlePheAlaLeuIleIleValGluThrLeuAlaTrpPh 164
  ::||| ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52513 GTGAGGTGGATCTATGGATTAGAAAGTGGTGGTGGCAGTGGGATTATGTT 52562

164 eThrValValAlaSerLeuPheAlaLeuProGlnMetArgGlyTyrg 181
  ||| ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52563 TTCCTTTTGGCCATTTTTCATCTCTCTATAGCAATTCAGACGTTTTTATC 52612

181 lnArgLeuAlaLysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGly 197
  || ::::: ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52613 AAACACACTCTTACTGGTTTGATAGATTTCTAGGGGCAGCTTTGCTATTT 52662

198 PheGlyIleHisLeuIleIle 204
  ||| ::::: ||| ::::: ||| ::::: ||| ::::: |||
52663 TTCGCCATTGCAATTATCATC 52683
```

OM of: US-09-466-935-4 to: N_Geneseq_0401.* out_format : pfs
Date: May 6, 2001 3:43 PM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

-MODEL=framet_p2n.model -DEV=xlp
-O=/cgn2_1/USPTO.spool/US09466935/runat_03052001_075710_14843/app_query.fasta_1.529
-DB=N_Geneseq_0401 -QFMT=fastap -SUFFIX=rng -GAPOP=12.000
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000 -LOOPEXT=0.000
-QGAPOP=4.500 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500
-DELOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blotsum62
-TRANS=human40 cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs
-NORM=ext -MINLEN=0 -MAXLEN=2000000000
-USER=US09466935_@CGN1_1_213 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPPY
-WAIT -THREADS=1

Search information block:

Query: US-09-466-935-4
Query length: 206
Database: N_Geneseq_0401.*
Database sequences: 678276
Database length: 291890651
Search time (sec): 172.640000

score_list:

Sequence	Strd Orig	zScore	Escore	Len	Documentation
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48443	+	1054.00	2252.72	840	1.5e-117
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52688	+	261.50	544.18	672	2.3e-22
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52689	+	190.00	390.36	639	8.4e-14
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A61501	+	179.50	357.24	1668	5.8e-12
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A61502	+	179.50	357.24	1668	5.8e-12
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A61502	+	122.00	236.69	1200	3.0e-05
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48442	+	122.00	236.41	1231	3.1e-05
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:T06767	+	111.50	219.58	720	0.0003
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:T06767	+	111.50	219.58	720	0.0003
/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:T96816	+	108.00	199.03	636	0.0038
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52691	+	103.00	202.58	588	0.0024
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52690	+	101.00	199.12	1063	0.0038
/SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:Q80215	+	96.50	182.95	401	0.0298
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C74616	+	94.50	189.25	1597	0.0133
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:T33622	+	90.50	165.57	1446	0.2773
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:C84375	+	90.00	162.75	1873	0.3980
/SIDS2/gcgdata/geneseq/geneseq/NA1995.DAT:Q80220	+	90.00	161.88	2029	0.4450
/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:T58559	+	90.00	161.00	2199	0.4980
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X87701	+	90.00	161.00	2199	0.4980
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X04732	+	90.00	161.00	2199	0.4980
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:D00787	+	90.00	161.00	2199	0.4980
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A00722	+	90.00	161.00	2199	0.4980
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:T33536	+	89.00	137.76	15239	9.81
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:T33219	+	86.50	159.09	1310	0.6366
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:T99070	+	86.50	159.09	1310	0.6366
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:T33220	+	86.50	152.47	2404	1.49
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:T99071	+	86.50	152.47	2404	1.49
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:Q04775	+	86.00	153.87	1915	1.24
/SIDS2/gcgdata/geneseq/geneseq/NA1996.DAT:T35230	+	85.50	144.98	3923	3.89
/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:T87858	+	85.00	155.47	1356	1.01
/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:T94571	+	85.00	155.47	1356	1.01
/SIDS2/gcgdata/geneseq/geneseq/NA1997.DAT:T74475	+	85.00	125.00	22243	50.38
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:T74486	+	84.00	133.53	8339	16.88
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:T212314	+	82.50	152.64	1071	1.45
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C50050	+	82.00	148.45	1425	2.49
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:T255887	+	82.00	106.24	68750	559.28
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C39157	+	81.50	144.32	2114	4.96
/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C45934	+	81.50	143.08	2114	4.96
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:X06875	+	81.50	137.46	3540	10.19

/SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:C76432 - 81.00 150.33 1.96 984
/SIDS2/gcgdata/geneseq/geneseq/NA2001.DAT:F4688 - 81.00 135.07 13.85 3993
/SIDS2/gcgdata/geneseq/geneseq/NA1998.DAT:V52152 + 80.50 131.94 20.69 4820
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:T212312 + 80.00 147.25 2.91 1071
/SIDS2/gcgdata/geneseq/geneseq/NA1999.DAT:T212313 + 80.00 147.25 2.91 1071

seq_name: /SIDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A48443

seq_documentation_block:

ID A48443 standard; DNA; 840 BP.
XX
AC A48443;
XX
DT 08-SEP-2000 (first entry)
XX
DE E. coli L-threonine resistance gene, rhtC.
XX
KW L-threonine resistance; L-threonine synthesis; rhtC;
KW L-homoserine; L-valine; L-leucine; ds.
XX
OS Escherichia coli.

Key Location/Qualifiers

CDS 187..807

FT /*tag= a "rhtC"

FT /product= "rhtC"

XX EP1013765-A1.

XX 28-JUN-2000.

XX 20-DEC-1999; 99EP-0125406.

XX 23-DEC-1998; 98RU-0123511.

XX (AJIN) AJINOMOTO KK.

XX Livshits VA, Zakataeva NP, Aleshin VV, Belareva AV, Tokhmakova IL;

XX WPI; 2000-414602/36.

XX P-PSDB; Y99598.

XX Novel Escherichia bacterium having enhanced L-threonine resistance due

XX to enhanced RhtC protein activity, used to produce L-threonine,

XX L-homoserine, L-valine and L-leucine -

XX Claim 9; Page 14-15; 24pp; English.

XX The present sequence is the L-threonine resistance gene, rhtC, from

XX Escherichia coli. This sequence may be used to impart L-threonine

XX resistance on E. coli bacteria, which would be useful for producing

XX a high yield of L-threonine. L-threonine resistance means that the

XX bacteria will be able to grow on a minimal medium containing

XX L-threonine at a concentration at which the corresponding wild-type

XX strain would not grow. Since the transformed bacteria can grow on the

XX minimal medium, it can synthesize L-threonine, which accumulates. The

XX accumulated amino acids can then be removed from the culture medium.

XX The bacterium of the present invention may also be used to synthesize

XX L-homoserine, L-valine and L-leucine at increased levels.

XX Sequence 840 BP; 165 A; 192 C; 248 G; 235 T; 0 other;

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

XX

```
1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
|||||
187 ATGTTGATGTTATTTCTACCGTCGCATGGTGCACATTTGGCGCTTAT 236
|||||
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
|||||
237 GAGCCCGGTCGCCGATTTCTTTTCTCTCAGACCGCTGTCAGTCGT 286
|||||
34 erArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
|||||
287 CCGCTAAAGAAGGATGATGGCGTCTGGGCATTACCTCGCGGTAATG 336
|||||
51 ValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleGluLysMe 67
|||||
337 GTTTGGCTGGGATGGCTGCTTGGCTGCAITTTGATTATCGAAAAAT 386
|||||
67 tAlaTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuTyrLeuCys 84
|||||
387 GGCCTGGCTGCATACGCTGATTATGTTGGCGGTGGCTGTATCTCTGCT 436
|||||
84 rPheMetGlyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSer 100
|||||
437 GGAATGGGTTACCAAGATGCTACGTTGGTGCACAGAAAAAGAGCGTTCT 486
|||||
101 AlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeu 117
|||||
487 GCACCTGGCCACAGTCGAGCTGGCGAAAGTGGCGCAGTTCTCTGAA 536
|||||
117 sGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGly 134
|||||
537 AGGTTTACTGACCAATCTCGCTAATCGAAAGCGATTATCTACTTTGGCT 586
|||||
134 erValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTyr 150
|||||
587 CGGTGTTCTCATTTGTCGGTGATAACGTTGGCACACCGCGCGCTGG 636
|||||
151 GlyIlePheAlaLeuIleValGluThrLeuAlaTrpPheThrValVa 167
|||||
637 GGCATTTTTCGGCTGATCATGTCGAAACGCTGGCGTGGTTTACCGTCGT 686
|||||
167 lAlaSerLeuPheAlaLeuProGlnMetArgGlyTyrGlnArgLeuA 184
|||||
687 TGGCAGCGCTGTTGCCCTGCGCAAAATGCGCGCTGTTATCAACGCTCG 736
|||||
184 lAlaSerTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
|||||
737 CGAAGTGGATTGATGGTTTTGCCGGGGCGGTATTTCGCCGATTGGCAIT 786
|||||
201 HisLeuIleIleSerArg 206
|||||
787 CATTTGATTATTCGCGG 804
|||||
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A52688

```
seq_documentation_block:
ID A52688 standard; DNA; 672 BP.
XX
AC A52688;
XX
XX
DT 03-JAN-2001 (first entry)
XX
DE Escherichia coli yahn gene.
XX
KW E. coli; yahn gene; amino acid production; excretion protein gene;
KW amino acid excretion protein; ds.
XX
OS Escherichia coli.
XX
XX
FH Key Location/Qualifiers
FT CDS 1..672
FT /*tag= a
FT /product= "yahn"
XX
```

```
PN EP1016710-A2.
XX
PD 05-JUL-2000.
XX
PF 17-DEC-1999; 99EP-0125263.
XX
PR 30-DEC-1998; 98RU-0124016.
PR 09-MAR-1999; 99RU-0104431.
XX
PA (AJIN ) AJINOMOTO CO INC.
XX
XX Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
PI Tokhmakova IL;
XX
XX WPI; 2000-414802/36.
DR P-PSDB; B01786.
XX
XX Increased production of L-amino acids by an Escherichia bacterium
PT comprises increasing the expression amount of an L-amino acid excretion
PT protein -
XX
XX Disclosure; Page 17-18; 29pp; English.
XX
XX The present sequence is the yahn gene (an excretion protein gene) of
CC Escherichia coli. The amino acid excretion protein produced from this
CC gene is involved in the production of amino acids, and an increase in its
CC expression leads to an increased accumulation of amino acids in the cell.
CC In this case, an increase in lysine, glutamic acid and proline is
CC achieved if multiple copies of the gene are transfected into a bacterium.
CC The bacterium used is E. coli.
XX
XX Sequence 672 BP; 130 A; 154 C; 178 G; 210 T; 0 other;
SQ

alignment_scores:
Quality: 261.50 Length: 206
Ratio: 1.937 Gaps: 3
Percent Similarity: 65.534 Percent Identity: 28.155

alignment_block:
US-09-466-935-4 x A52688 ..

Align seg 1/1 to: A52688 from: 1 to: 672

4 LeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerProgl 20
|||||
58 GTTACCTGACCGTAGGACTG...TTCGTGATTACTTTTTTATCCGG 104
|||||

20 yProAspPhePheValSerGlnThrAlaValSerArgSerArgLysG 37
|||||
105 AGCCAATCTCTTGTGGTAGTACAAACAGCGCTGGCTTCCGGTCGACGCG 154
|||||

37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
|||||
155 CAGGGTGCTGACCGGCTGGCGGTGGCGATGCGCATTTTATTTCC 204
|||||

54 GlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMetAlaTrpLe 70
|||||
205 GGGTTGGGTTGTTGGTCTTGCACACGCTAATTACGAGTGTGAGGAGAT 254
|||||

70 uHisThrLeuIleMetValGlyGlyLeuTyrLeuCysTrpMetGlyT 87
|||||
255 TTTTTCGCTTATCAGAATCGTCGCGGCGCTTATCTCTTATGTTGGGT 304
|||||

87 yTrpMetLeuArgGlyAlaLeuLysLysGluAlaValSerAlaProAla 103
|||||
305 GGTGAGCATGCGC.....CGCCAGTCAACA 330
|||||

104 ProGlnValGluLeuAlaLysSerGlyArgSer..... 114
|||||
331 CCGCAATGAGCAGCTACACACCGATTAGCGCCCTCGTATGCTTT 380
|||||

115 .PheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIle 131
|||||
```

```

131 TTTTGGCGCGGATTAATACCGATCTCTAAACCGCAACCGTTTAT 430
131 yrPheGlySerValPheSerLeuPheValGlyAspAsnValGlyThrThr 147
131 TTTTATCAGTATTTCTCAGTACATTAATGCCGAACACCAACATCG 480
148 AlaArgTrpGlyIlePheAlaLeuIleValGluThrLeuAlaTrpPh 164
481 GCACGTTTAAATGGCTGGCGGGATTTGCTGCGATCAATTATCTGGG 530
164 eThrValValAlaSerLeuPheAlaLeuProGlnMetArgGlyTyrg 181
531 AGTTTCTTCTAGTCAGCGCTTCTTTCGCCGCTGCTGCGCTATG 580
181 InArgLeuAlaLeuTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGly 197
581 GCGGTATGCAACGGGTTGCCAGTCGGGTTATTGTCGCAATTATTGGTGA 630
198 PheGlyIleHisLeuIle 203
631 TTCGGCTACCGCTGATT 648

```

seq_name: /SDS2/gcdata/geneseq/geneseq/NA2000.DAT:A52689

```

seq_documentation_block:
ID_ A52689 standard; DNA; 639 BP.
AC_ A52689;
DT_ 03-JAN-2001 (first entry)
DE_ Escherichia coli yeas gene.
KW_ E. coli; yeas gene; amino acid production; excretion protein gene;
KW_ amino acid excretion protein; ds.
OS_ Escherichia coli.
FH_ Key Location/Qualifiers
FT_ CDS 1..639
FT_ /*tag= a
FT_ /product= "Yeast"
XX_ EP1016710-A2.
XX_
XX_ 05-JUL-2000.
XX_
XX_ 17-DEC-1999; 99EP-0125263.
XX_
XX_ 30-DEC-1998; 98RU-0124016.
XX_
XX_ 09-MAR-1999; 99RU-0104431.
XX_
XX_ (AJIN ) AJINOMOTO CO INC.
XX_
XX_ Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
XX_ Tokhmakova IL;
XX_
XX_ WPI; 2000-414802/36.
XX_
XX_ P-PSDB; B01787.
XX_
XX_ Increased production of L-amino acids by an Escherichia bacterium
XX_ comprises increasing the expression amount of an L-amino acid excretion
XX_ protein -
XX_
XX_ Disclosure; Page 19-20; 29pp; English.
XX_
XX_ The present sequence is the yeast gene (an excretion protein gene) of
XX_ Escherichia coli. The amino acid excretion protein produced from this
XX_ gene is involved in the production of amino acids, and an increase in its
XX_ expression leads to an increased accumulation of amino acids in the cell.
XX_ In this case, an increase in lysine, alanine, valine, histidine,
XX_ isoleucine, glutamic acid and proline is achieved if multiple copies of

```

```

CC_ the gene are transfected into a bacterium. The bacterium used is E. coli.
XX_
SQ_ Sequence 639 BP; 142 A; 128 C; 161 G; 208 T; 0 other;

alignment_scores:
Quality: 190.00 Length: 202
Ratio: 1.624 Gaps: 6
Percent Similarity: 57.921 Percent Identity: 28.218

alignment_block:
US-09-466-935-4 x A52689 ..
Align seg 1/1 to: A52689 from: 1 to: 639

13 ILeValAlaLeuMetSerProGlyProAspPhePhePheValSerGlnTh 29
131 TTTTGGCGCGGATTAATACCGATCTCTAAACCGCAACCGTTTAT 430
52 ATTTTATTGTGTGTGGTGGCGGCAATACCCCTGTTGTACTCAAAA 101
29 rAlaValSerArgSerArgLysGluAlaMetMetGlyValLeuGlyIle 46
102 TAGGTCAGTACGGGTATGAAGCGGTTATCTTGGCGCTCGGTGTAT 151
46 hrCysGly.....ValMetValTrpAlaGlyIleAlaLeu 57
152 TTATTGGCGATGCGGTATTGATGTTCTGGCATGGCTGGAGTGGCGACA 201
58 LeuGlyLeuHisLeuIleIleGluLysMetAlaTrpLeuHisThrLeuI 74
202 TTA.....ATTAAGACCCCGCATATTATTCAACATGT 236
74 eMetValGlyGlyGlyLeuTyriLeuCysTrpMetGlyTyriGlnMetLeu 91
237 ACGTATCTTGTGGTGGCTTTTATTGCTCTATCTGGGAGTAAATTCCTT 286
91 rgGlyAlaLeuLysLysGluAlaValSerAlaProAlaProGlnValGlu 107
287 ACGGACCTTGAAGGTAAATAATAGCGAGCCCAATCCGATGAGCCCAA 336
108 LeuAlaLysSerGlyArgSerPheLeuLysGlyLeuLeuThrAsnLeuAl 124
337 TAC.....GGTGCTATTTTAAACGCGGTTAATTTTGGCGCTGAC 377
124 aAsnProLysAlaIleIleTyriPheGlySerValPheSerLeuPheValG 141
378 TAATCCGAAGACCATTTGTCTATGTCTGCTTTTCGTACAGTTTATCG 427
141 lyAspAsnValGlyThrAlaArgTrpGlyIlePheAlaLeuIleIle 157
428 ATGTTAATGCCCCACATACG.....GGAATTTTCATTCTTTATTCTG 468
158 ValGluThrLeuAlaTrp.....PheThrValVal 167
469 GCGGCGACGCTGGAACCTGCTGCTATTTGAGCTTCTCTGATTAT 518
167 lAlaSerLeuPheAlaLeuProGlnMetArgGlyTyriGlnArgLeuAl 184
519 ATCTGTGTGCTTTT...GTCCAGCAGTACATACGTACCAAAAAGAACTGG 565
184 lAlysTrpIleAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
566 CTAAGTTGGCAACTCACTGATTGGTTTGTGATGTTCTGTTGCTGCTGCC 615
201 HisLeu 202
616 CGACTG 621

```

```

seq_name: /SDS2/gcdata/geneseq/geneseq/NA2000.DAT:A61501
seq_documentation_block:
ID_ A61501 standard; DNA; 1668 BP.
XX_
AC_ A61501;

```



```

587 CTGACATCGATCATTTAAACGCTGTCGCCAGGCTCTGGTGCATCAACAC 636
26 lserGlnThrAlaValSerArgSerArgLysGluAlaMetMetGlyValL 43
637 TATGACCACTCGCTCAACACGGTTATCCGGCGGT ..... 673
43 euGlyIleThrCysGlyVal.....Met 50
674 ..GGGCTCTATTGTGGCTCTAGACCGGACTGGCGATTATTCGTGT 721
51 ValTPrAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIle.....G 65
722 GGTGGCTGGGGTGGGACGCTATTTCCCGCTCAGTGATTGCGTTGA 771
65 uLysMetAlaTrpLeuHisThrLeuIleMetValGlyGlyLeuTyrL 82
772 AGTTGTAAGTGG.....GCAGCGCGGCTTACT 800
82 euCysTrpMetGlyTyrGlnMetLeuArg.....GlyAlaLeuLysLys 96
801 TGATTGGCTGGGAATCCAGCAGTGGCGCGCTGTGCAATTGACCTT 850
97 GluAlaValSerAlaProAlaProGlnValGluLeuAlaLysSerGlyAr 113
851 AAATCGCTGGGCTCTACTCAATCCGCTCGACATTG..... 886
113 gSerPheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAlaIle 130
887 ....TTCACGCGCGCAGTGTGTGTAATCTCACCAATCCCAAGTATTG 932
130 leTyrPheGlySerValPheSerLeuPheValGlyAspAsnValGlyThr 146
933 TGTTTCTGGCGGCGCTATTCGCAATTCATGCGCGCACGCCGCAA 982
147 ThrAlaArgTrpGlyIlePheAlaLeuIleIleValGluThrLeuAlaTr 163
983 CTGATGTCAGTATCGTCTCGGCTCACC..... 1012
163 pPheThrValValAlaSerLeuPheAlaLeuProGlnMetArgGlyT 180
1013 ....ACTATTGTGGTGCATATATTGTGATGATCGGTACGCCACCCTTG 1058
180 yrGlnArgLeuAlaLysTrpIleAspGly..... 189
1059 CTCACGCGATTGCTCTATGATTAAGACCAAGCAGATGAAGCGCTG 1108
190 .....PheAlaGlyAlaLeuPheAlaGlyPheGlyIleHisLeuIle 204
1109 AATAAGATTTTCGCTCGTCTTTTATGCTGGTGGGCGCTGTACATC 1158
204 eSerArg 206
1159 GCGGAGG 1165
```

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1996.DAT:T06767
seq_documentation_block:
XX T06767 standard; DNA; 720 BP.
XX AC T06767;
XX DT 01-JUL-1996 (first entry)
XX DE Mel-linked mlga gene.
XX KW Marine mela; selectable marker; oyster larva settlement;
XX KW pigmentation; melanin; p-hydroxyphenylpyruvate-hydroxylase;
XX KW mlga gene; ss.
XX OS Shewanella colwelliana strain lSTDYF.
XX Key Location/Qualifiers
FH
```

```

FT RBS 154..159
FT FT /*tag= a
XX US5474933-A.
XX PD 12-DEC-1995.
XX PF 21-MAR-1990; 90US-0496804.
XX PR 08-NOV-1993; 93US-0148945.
XX PR 21-MAR-1990; 90US-0496804.
XX PR 10-NOV-1992; 92US-0974837.
XX (UYMA-) UNIV MARYLAND BALTIMORE.
XX Fuqua WC, Weiner RM;
XX WPI; 1996-039515/04.
XX P-PSDB; R87527.
XX Novel gene encoding marine mela from Shewanella - useful as
XX selectable marker in genetic engineering and for inducing larval
XX oyster settlement
XX Example 7; Fig 14; 47pp; English.
XX The 5' end of an open reading frame (T06767) was identified
XX directly downstream of the Shewanella colwelliana mela gene
XX (T06766), and was designated mlga (mel-linked gene). The
XX role of the encoded protein (R87527) was unclear as deletion
XX subcloning in E. coli demonstrated that only mela was required
XX for melanogenesis.
XX Sequence 720 BP; 180 A; 165 C; 165 G; 210 T; 0 other;

alignment_scores:
Quality: 111.50 Length: 203
Ratio: 0.987 Gaps: 7
Percent Similarity: 55.665 Percent Identity: 25.123

alignment_block:
US-09-466-935-4 x T06767 ..
Align seg 1/1 to: T06767 from: 1 to: 720

21 ProAspPhePhePheValSerGlnThr..... 29
128 CCGACCTTTTTCGTCTCAATTACACCAAGTATGTGTATGACACTGGC 177
30 .....AlaValSerArgSerArgLysGluAlaMetMetGlyValL 43
178 CATGACTCTCGGTATGAGTATGCGTGTGCGCGCAACCTTATGGATGATGG 227
43 euGlyIleThrCysGlyValMetValTrpAlaGlyIleAlaLeuLeuGly 59
228 TTGCTGAGCTAGCAGCGTTCCTCGTGGCGGATTCGCGCGTAAATGGGT 277
60 LeuHisLeuIleLeuLysMetAlaTrpLeuHisThrLeuIleMetVa 76
278 GTCGCCAGTATGATGCTCAACTATCCACAATCTTCGATATTTAAATG 327
76 lGlyGlyGlyLeuTyrLeuCysTrpMetGlyTyrGlnMetLeuArgGlyA 93
328 GGTGCGTGGGCTCTATCTTGGTTACATCGGCATTAGCATGTGGCGGCCA 377
93 lalLeuLysLysGluAlaValSerAlaProAlaProGlnValGluLeuAla 109
378 AAGGAAATGGCCAACTTGACATAACCTCCAGTCAGATC..... 418
110 LysSerGlyArgSerPheLeu...LysGlyLeuLeuThrAsnLeuAlaAs 125
419 ...AGTAATCGAGCGCTAATAACTCAAGGCTTTGTCCACCGCAATTGCTAA 465
```


ID T96816 standard: DNA; 2374 BP.
 XX T96816;
 XX
 XX 12-MAR-1998 (first entry)
 XX
 XX DNA encoding LysG, LysE and ORF3 from *Corynebacterium glutamicum*.
 XX
 XX LysG; LysE; ORF3; lysine transport; regulatory protein; export protein;
 KW Microbial production; amino acid; animal feed additive; ds.
 XX
 XX *Corynebacterium glutamicum*.
 OS
 FH Key Location/Qualifiers
 FT CDS complement (82..954)
 FT /*tag= a
 FT /label= LysG
 FT 1016..1726
 FT /*tag= b
 FT /label= LysE
 FT complement (1723..2373)
 FT /*tag= c
 FT /label= orf3
 XX
 XX DE19548222-AL.
 XX
 XX 26-JUN-1997.
 XX
 XX 22-DEC-1995; 95DE-1048222.
 XX
 XX 22-DEC-1995; 95DE-1048222.
 XX
 XX (KBRJ) FORSCHUNGSZENTRUM JUELICH GMBH.
 XX
 XX Eggeling L, Sahm H, Vrije M;
 XX
 XX WPI: 1997-333867/31.
 DR P-FSDB; W3714-16.
 XX
 PT Increasing microbial production of amino acids, especially lysine -
 PT by improving export carrier activity or corresponding gene
 PT expression, also new export and regulatory genes from
 PT *Corynebacterium*
 XX
 PS Claim 23 and 26; Page -; 16pp; German.
 XX
 CC This DNA, isolated from *Corynebacterium glutamicum*, contains the LysG,
 CC LysE and ORF3 genes. LysG and LysE encode a lysine transport regulatory
 CC protein and an export protein, respectively. Microbial production of
 CC amino acids (A) is improved by increasing the export-carrier activity
 CC and/or the export gene expression in a microorganism that produces (A).
 CC The method is specifically used to increase production of lysine,
 CC used as an animal feed additive. Other (A) are variously useful as
 CC pharmaceuticals, condiments and intermediates for fine chemicals.
 CC This method increases the amount of (A) secreted into the culture medium.
 CC Export of (A) has been found to depend on a single gene.
 CC NB. This sequence has been created from the information given in table 2
 CC of the specification.
 XX
 SQ Sequence 2374 BP; 526 A; 640 C; 648 G; 560 T; 0 other;

alignment_scores:
 Quality: 108.00 Length: 263
 Ratio: 0.900 Gaps: 13
 Percent Similarity: 45.627 Percent Identity: 23.574

alignment_block:
 US-09-466-935-4 x T96816 ..

Align seg 1/1 to: T96816 from: 1 to: 2374

1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17

```

1022 ATCATGGAAATCTTCATTACAGGTCTGCTTTTGGGGCCAGCTTTTACT 1071
17 tSerProGlyProaspPhePheValSerGlnThrAlaValSerArgS 34
1072 GTCCATCGGACCGCAGAGTACTGGTGGTAAACAAGGAATTAAGCGC. 1120
34 exArgLysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMet 50
1121 .....GAGGACTCATTCGGGTCTCTCGGTGGTAAATTCTGAC 1162
51 ValTrpAlaGlyIleAla...LeuLeuGlyLeuHisLeu..... 62
1163 GTCTTTTGTTCATCGCGGCACCTTGGGGCTTGATCTTTTGTCCAATGC 1212
63 .....IleIleGluLysMetAlaTrpLeuHisThrLeuIleMetV 76
1213 CGGCGCATCGTCTCGATATATGCGCTGG..... 1243
76 alGlyGlyLeuTrpLeuCysTrpMetGlyTyrGlnMetLeuArgGly 92
1244 ..GGTGGCATCGCTTACCTGTTATGGTTGCGGTCATGGCAGCAAGAC 1291
93 AlaLeuLysLysGluAlaValSerAlaProAlaProGlnVal..... 106
1292 GCCATGACAAACAAG.....GTGGAAGCGCCACAGATCATTTGAAGA 1332
107 .....GluLeuA 109
1333 AACAGAACCAACCGTCCCGATGACACGCTTTGGGCGGTTCGCGGTGG 1382
109 laLysSerGlyArgSer..... 114
1383 CCACGTGACACGCGCAACCGGGTGGAGGTGAGCGTGCATAAGCAG 1432
115 .....PheLeuLysGlyLeuLeuThrAsnLeuAlaAs 125
1433 CGGGTTGGGTAACCCCATGTTGATGGCAATCGTGTGACCTGGTTGAA 1482
125 nProLys.....AlaIleIleTyrPheGlySerValPheSerL 138
1483 CCGGATCGTATTTGGACGGGTTGTGTTTATCGCGGC..... 1522
138 euPheValGlyAspAsnValGlyThrThrAlaArgTrpGlyIlePheAla 154
1523 .....GTGCGCGCAATACGCGCACACGCGAGCGTGGATTTTCGCGCT 1567
155 LeuIleIleValGluThrLeuAlaTrpPheThrValValAlaSerLeuPh 171
1568 GGGCGCTTCGCGCAAGCCTGATCTGGTCCGCTGGTGGG.....TT 1610
171 e.....AlaL 173
1611 TCGGCGCAGCAGCATTTGTCACGCCCGCTGTCCAGCCCAAGGTGTGCGCG 1660
173 euProGlnMetArgArgGlyTyr.....GlnArgLeuAlaLysTrpIle 187
1661 TGGATCAAGCTCGTGGCAGTGTGTGATGACCGCATTCGCGCATCAAACT 1710
188 AspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIle 200
1711 GAT...GTTGATGGGTAGTTTTCGCGGTTTGAATC 1746

```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA2000.DAT:A52691

seq_documentation_block:
 ID A52691 standard; DNA; 636 BP.

XX AC A52691;

XX DT 03-JAN-2001 (first entry)
 XX DE Escherichia coli yggA gene.

```

XX E. coli; yggA gene; amino acid production; excretion protein gene;
KW amino acid excretion protein; ds.
XX
OS Escherichia coli.
XX
FH Key Location/Qualifiers
FT CDS 1..636
FT /*tag= a
FT /product= "YggA"
XX
XX EP1016710-A2.
XX
XX PD 05-JUL-2000.
XX
XX PF 17-DEC-1999; 99EP-0125263.
XX
XX PR 30-DEC-1998; 98RU-0124016.
XX
XX PR 09-MAR-1999; 99RU-0104431.
XX
XX PA (AJIN ) AJINOMOTO CO INC.
XX
XX PI Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
XX PI Tokhmakova IL;
XX
XX DR WPI; 2000-414802/36.
XX
XX DR P-PSDB; B01789.
XX
XX PT Increased production of L-amino acids by an Escherichia bacterium
XX PT comprises increasing the expression amount of an L-amino acid excretion
XX PT protein -
XX
XX PS Disclosure; Page 24; 29pp; English.
XX
XX CC The present sequence is the yggA gene (an excretion protein gene) of
XX CC Escherichia coli. The amino acid excretion protein produced from this
XX CC gene is involved in the production of amino acids, and an increase in its
XX CC expression leads to an increased accumulation of amino acids in the cell.
XX CC In this case, an increase in arginine, glutamic acid and lysine is
XX CC achieved if multiple copies of the gene are transfected into a bacterium.
XX CC The bacterium used is E. coli.
XX
XX SQ Sequence 636 BP; 120 A; 143 C; 180 G; 193 T; 0 other;

alignment_scores:
  Quality: 103.00 Length: 194
  Ratio: 1.030 Gaps: 9
Percent Similarity: 51.546 Percent Identity: 27.835

alignment_block:
US-09-466-935-4 x A52691 ..

Align seg 1/1 to: A52691 from: 1 to: 636

4 LeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerProG1 20
   ||| ||| ::||| |||
25 CTTCGACTGGGGCGGCTATGATC.....CTACCGCTCGG 59
   |||::: |||::: |||::: |||:::
20 YProAspPhePhePheValSerGlnThrAlaValSerArgSerArgLysG 37
   |||::: |||::: |||::: |||:::
60 TCACAAATGCTTTGTGATGATCAGGCATAGCTGCTCAGTACCACA 109
   |||::: |||::: |||::: |||:::
37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
   |||::: |||::: |||::: |||:::
110 TTATGATTCCTTACCTTGTGTCATACGAGATTGGTCTCTGATTCGCC 159
   |||::: |||::: |||::: |||:::
54 GlyIleAlaLeuLeuGlyLeuHisLeuIleGluLysMetAlaTrpLe 70
   ||| ||| ::||| |||
160 GGGATT.....TTTGGTGGCAGCGGTATTGATGATCAGTCGCGGTGTT 203
   |||::: |||::: |||::: |||:::
70 uHisThrLeuIleMetValGlyCysGlyLeuTyrrLeuCysTrpMetGlyT 87
   | ::|||::: ||| ||| |||

```

```

204 GCTGGCGCTGGTCCACCTGGGGCGGCTAGCCTTCTTCTGCTGGTATGGTT 253
87 yrrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSerAlaProAla 103
   :: ::::: |||::: |||
254 TTGGCGCTTTTAAACAGCAATGAGCAGTAATATTGAGTTAGCCAGCGCC 303
   |||::: |||::: |||::: |||:::
104 ProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuLe 120
   |||::: |||::: |||::: |||:::
304 .....GAAGTCATGAAGCAAGGCAGA.....TGGAAATATATCGC 338
120 uThrAsnLeuAla.....AsnProLysAlaIleIleTyrrPheG 133
   ||| ||| ||| ||| ||| |||
339 CACCATGTTGGCAGTACCTGCTGCTGATCCGATGTTTACCTGGATACTT 388
133 lySerValPheSerLeuPheValGlyAspAsnValGlyThrThrAlaArg 149
   |||::: |||::: |||::: |||:::
389 TTGTTGTTACTGGGCGGCACTTGGCGGCAACTTGTATGTGGAACCAACGC 438
150 TrpGlyIlePheAlaLeu...IleIleValGluThrLeuAlaTrpPheTh 165
   ||| ||| ||| ||| ||| |||
439 TGG.....TTTGCACTCGGGCAATTAGCGCTCTTCTCTGTGTTCTT 482
165 rValValAlaSerLeuPheAla.....LeuProGlnMetArgArgGly. 179
   ::||| ||| ||| ||| |||::: |||:::
483 TGGTCTGGCTCTTCTCGACCTGGCTGGCACCCTGCTCTGCGACGGCAA 532
180 ..TyrGlnArgLeuAlaLysTrpIleAspGly 189
   ||| |||::: |||::: |||
533 AAGCACAGCGCATATCAATCTGGTTGTGGGA 564

seq_name: /SDS2/gcgdata/geneseq/geneseq/NA2000.DAT:A52690
seq_documentation_block:
ID A52690 standard; DNA; 588 BP.
XX AC A52690;
XX DT 03-JAN-2001 (first entry)
XX DE Escherichia coli yfik gene.
XX KW E. coli; yfik gene; amino acid production; excretion protein gene;
XX KW amino acid excretion protein; ds.
XX OS Escherichia coli.
XX FH Key Location/Qualifiers
XX FT CDS 1..588
XX FT /*tag= a
XX FT /product= "yfiK"
XX
XX PN EP1016710-A2.
XX
XX PD 05-JUL-2000.
XX
XX PF 17-DEC-1999; 99EP-0125263.
XX
XX PR 30-DEC-1998; 98RU-0124016.
XX
XX PR 09-MAR-1999; 99RU-0104431.
XX
XX PA (AJIN ) AJINOMOTO CO INC.
XX
XX PI Livshits VA, Zakataeva NP, Nakanishi K, Aleshin VV, Troshin PV;
XX PI Tokhmakova IL;
XX
XX DR WPI; 2000-414802/36.
XX
XX DR P-PSDB; B01788.
XX
XX PT Increased production of L-amino acids by an Escherichia bacterium
XX PT comprises increasing the expression amount of an L-amino acid excretion
XX PT protein -
XX
XX PS Disclosure; Page 22; 29pp; English.

```

The present sequence is the yfiK gene (an excretion protein gene) of *Escherichia coli*. The amino acid excretion protein produced from this gene is involved in the production of amino acids, and an increase in its expression leads to an increased accumulation of amino acids in the cell. In this case, an increase in threonine, histidine, glutamic acid and proline is achieved if multiple copies of the gene are transfected into a bacterium. The bacterium used is *E. coli*.

Sequence 588 BP; 113 A; 134 C; 167 G; 174 T; 0 other;


```

245 CGCTTGCAGTGTGTGCTGCTATGCGCTGTGGCAATGTGTTT..... 288
174 ProGlnMetArgArgGlyTyr.....GlnArgLeuAlaLysTrpI 187
289 .....CGTCGTGAGTGGTGCAGTCGCCAGTGTGCAGACTGGCT 329
187 eAspGlyPheAlaGlyAlaLeuPheAlaGlyPheGlyIleHisLeuIleI 204
330 CGGACGCGAGTTTGGCCAGCGCTTGGCCGGGCTGGGGTTGAACCTGGCGT 379
204 leSerArg 206
380 TTGCGCAG 387
```

seq_name: /SIDS2/gcgdata/geneseq/geneseqn/NA1999.DAT:233622

seq_documentation_block:

ID Z33622 standard; cDNA; 1597 BP.

XX Z33622;

DT 08-DEC-1999 (first entry)

DE Human breast tumour-associated EST 12.

XX Expressed sequence tag; EST; human; breast; cancer; gene therapy;
KW treatment; tumour; cytostatic; medicament; ss.

OS Homo sapiens.

XX DE19813839-Al.

XX PD 23-SEP-1999.

XX PF 20-MAR-1998; 98DE-1013839.

XX PR 20-MAR-1998; 98DE-1013839.

XX PA (META-) METAGEN GES GENOMFORSCHUNG MBH.

XX PI Specht T, Hinzmann B, Schmitt A, Pilarsky C, Dahl E, Rosentahl A;

XX DR WPI; 1999-528981/45.

XX Human nucleic acid sequences and protein products from tumor breast
tissue, useful for breast cancer therapy -

XX PS Claim 3; 93; 188pp; German.

XX This invention describes novel human nucleic acid sequences from tumor
breast tissue which have cytotatic activity. The nucleic acid sequences
can be used to produce and isolate full-length gene sequences. They can
be used to express proteins, which can be used as tools to find an
activity against breast cancer. The sequences can be used in sense or
antisense form. They are especially useful for medicaments for gene
therapy to treat breast cancer. 233611-248617 represents expressed
sequence tags described in the method of the invention.

XX SQ Sequence 1597 BP; 476 A; 229 C; 337 G; 555 T; 0 other;

alignment_scores:

Quality: 90.50 Length: 180
Ratio: 1.077 Gaps: 11
Percent Similarity: 46.567 Percent Identity: 27.778

alignment_block:

US-09-466-935-4 x Z33622 ..

Align seg 1/1 to: Z33622 from: 1 to: 1597

53 AlaGlyIleAlaLeuLeuGlyLeuHisLeuIleGluLysMetAlaTr 69

```

29 GCGGCGTGCCTCTCTG..... 111
69 pLeuHisThrLeuIleMetValGlyGlyLeuTyrLeuCysTrpMetG 86
52 G.....ATGGTGATATGGGTGATGCCCGCACAAATTCACGGATG 92
86 lyTyrGlnMetLeuArgGlyAlaLeuLysGluAlaValSerAla... 101
93 GCCTAGCAAT.....GGTGCCTCTTTTACTGAGGCTTATCAAGTGT 136
102 .....ProAlaProGln..... 105
137 TTAAGTACTCTGTGCTGTCTTCTGTCATGAGTTGCCTCATGAATTAGG 186
106 .....ValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuL 120
187 TGACTTGTCTTCTACTAAGGCTGCATGACCGTTAAGCAGGCTGCC 236
120 euThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGlySerValPhe 136
237 TTTATAAT...GCATTGTCAGCCATGCTGCCGTATCTTGAATGGCAACA 283
137 SerLeuPheValGlyAspAsnValGlyThrAlaArgTrpGlyIlePh 153
284 GGAATTTTCATTGCTCATTATGCTGAAATGTTCTATGTGG...ATATT 330
153 eAlaLeuIleIleValGluThrLeuAlaTrpPheThrValAlaSerL 170
331 TGCACCTT.....ACTGCTGCTTATTCATGTATGTGCTCTGG 368
170 euPheAlaLeuProGlnMet..... 176
369 TTGATATGTGTACCTGAAATGCTGCACAATGCTAGTGACCATGGATGT 418
177 ...ArgArgGlyTyrGlnArgLeuAlaLysTrpIleAspGlyPheAlaGl 192
419 AGCCGCTGGGGTATTTCTTTTACAGAAAT.....GCTGG 453
192 yAlaLeuPheAlaGlyPheGlyIleHisLeuIleIleSer 205
454 GATGCTTTTG...GGTTTGGAAATATGTACTATTATTC 490
```

OM of: US-09-466-935-4 to: Issued_Patents_NA: * out_format : pfs

Date: May 6, 2001 3:40 PM

About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:

```
-MODEL=framet-p2n.model  
-O/cgn2_1/USPTO.spool/US09466935/runat_03052001_075709_14814/app_query.fasta_1.529  
-DB-Issued_Patents_NA -OFT=fastap -SUFFIX=rni -GAPOP=12.000  
-GAPEXT=4.000 -MINMATCH=0.100 -LOOPEXT=0.000 -LOOPEXT=0.000  
-GAPOP=4.500 -GAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500  
-GAPOP=6.000 -GAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500  
-DELOP=6.000 -DELEXT=7.000 -START=1 -MATRIX=blosum62  
-TRANS-human40.cdi -LIST=45 -DOCALIGN=200 -THR_SCORE=pct  
-THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs  
-NORM=ext -MINLEN=0 -MAXLEN=2000000000  
-USER=US09466935_@cgn1_1_59 -NCPU=6 -ICPU=3 -LONGLOG -NO_XLPXY  
-WAIT -THREADS=1
```

Search information block:

```
Query: US-09-466-935-4  
Query length: 206  
Database: Issued_Patents_NA: *  
Database sequences: 302621  
Database length: 87301344  
Search time (sec): 80.320000
```

score_list:

Sequence	Strid	Orig	Zscore	EScore	Len	Documentation
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-476-254-6 +	111.50	238.89	1.0e-05	720	1	Patent
/cgn2_6/ptodata/2/ina/5A_COMB.seq:5474933-3 +	107.00	228.10	4.0e-05	720	1	Patent
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-847-743B-7 -	90.00	173.96	0.0419	2199	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-096-277-7 -	90.00	173.96	0.0419	2199	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-201-7 -	90.00	173.96	0.0419	2199	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-241-7 -	90.00	173.96	0.0419	2199	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-550-815-7 -	90.00	173.96	0.0419	2199	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-703-089-7 -	90.00	173.96	0.0419	2199	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCPT-US92-04295A-7 -	90.00	173.96	0.0419	2199	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-390-878-17 -	89.00	148.34	1.12	15239	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-311-023-1 +	86.50	171.78	0.0554	1310	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-311-023-3 +	86.50	164.50	0.1410	2404	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-395-246C-1 +	85.50	156.22	0.4074	3924	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-335-489-1 +	82.00	113.49	97.83	68750	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-940-424-10 +	79.00	154.56	0.5039	1229	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-881-857-1 -	79.00	151.73	0.7244	1556	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-233-342A-1 -	79.00	151.73	0.7244	1556	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-847-743B-12 -	78.50	147.47	1.25	2010	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-201-12 -	78.50	147.47	1.25	2010	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-330-161-10 -	78.50	147.47	1.25	2010	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-241-12 -	78.50	147.47	1.25	2010	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-440-401-10 -	78.50	147.47	1.25	2010	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-419-878B-10 -	78.50	147.47	1.25	2010	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCPT-US92-04295A-12 -	78.50	147.47	1.25	2010	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-096-277-6 -	78.50	146.58	1.40	2164	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-550-815-6 -	78.50	146.58	1.40	2164	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-703-089-6 -	78.50	146.58	1.40	2164	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-940-424-12 +	77.00	149.30	0.9898	1278	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-183-253-3 +	77.00	148.93	1.04	1318	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-09-183-253-1 +	77.00	139.53	3.47	2887	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-888-497-33 -	77.00	119.50	45.23	15328	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCPT-US94-07926-33 -	77.00	119.50	45.23	15328	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-612-734B-1 +	76.50	134.28	6.80	4047	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-612-734B-3 +	76.50	132.23	8.84	4800	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-728-214-13 +	76.50	131.27	10.00	5199	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCPT-US91-07046-1 +	75.00	131.70	9.46	3716	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-847-743B-25 -	74.00	134.40	6.69	2431	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-241-25 -	74.00	134.40	6.69	2431	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-241-25 -	74.00	134.40	6.69	2431	1	
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCPT-US92-04295A-25 -	74.00	134.40	6.69	2431	1	

```
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-07-847-743B-23 - 74.00 134.11 6.95 2490  
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-201-23 - 74.00 134.11 6.95 2490  
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-456-241-23 - 74.00 134.11 6.95 2490  
/cgn2_6/ptodata/2/ina/5A_COMB.seq:PCPT-US92-04295A-23 - 74.00 134.11 6.95 2352  
/cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-922-837-1 + 73.50 133.59 7.42
```

seq_name: /cgn2_6/ptodata/2/ina/5A_COMB.seq:US-08-476-254-6

```
seq_documentation_block:  
; Sequence 6, Application US/08476254  
; Patent No. 5846531  
; GENERAL INFORMATION:  
; APPLICANT: WEINER, RONALD M.  
; TITLE OF INVENTION: FUQUA, WILLIAM C.  
; NUMBER OF SEQUENCES: 11  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: WAYSON COLE STEVENS DAVIS, P.L.L.C.  
; STREET: 1400 K STREET NW  
; CITY: WASHINGTON  
; STATE: D.C.  
; COUNTRY: USA  
; ZIP: 20005-2477  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/476,254  
; FILING DATE: 06-JUN-1995  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: POULOS ILL. JAMES A.  
; REGISTRATION NUMBER: 31,714  
; REFERENCE/DOCKET NUMBER: JAP30319C  
; TELEPHONE: 202 628-0088  
; TELEFAX: 202 628-8034  
; INFORMATION FOR SEQ ID NO: 6:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 720 base pairs  
; TYPE: nucleic acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; MOLECULE TYPE: DNA (genomic)  
; HYPOTHETICAL: NO  
; ANTI-SENSE: NO  
; FEATURE:  
; NAME/KEY: CDS  
; LOCATION: 167..628  
; US-08-476-254-6
```

```
alignment_scores:  
Quality: 111.50 Length: 203  
Ratio: 0.987 Gaps: 7  
Percent Similarity: 55.665 Percent Identity: 25.123
```

alignment_block:

US-09-466-935-4 x US-08-476-254-6 ..

Align seg 1/1 to: US-08-476-254-6 from: 1 to: 720

```
21 ProAspPhePheValSerClnThr..... 29  
128 CCGACCTTTTTCGTCTCAATACACCGAGTGTGTGTGACACTGGC 177  
30 .....AlaValSerArgSerArgLysGluAlaMetGlyVal 43  
178 CATGACTCTCGGTATGATGCTGTGCGCGCAACCTTATGATGATGG 227  
43 euglytlethrcysglyvalmetvaltrpAlaGlyIleAlaLeuLeuGly 59
```



```

992      |||      ::|||:::
112 GlyArgSerPheLeuLysGlyLeuThrAsnLeuAlaAsnProLysAl 128
962 TTCAGACCG.....AAGGCTCTGCCGAAGAC 937
128 alleIleTyrPheGlySerValPheSerLeuPheValGlyAspAsnValG 145
936 GGTCAAG.....CAGCTTTTCGCTGTTCTTCTGGTTTTCAGTAG 896
145 ly.Thr.....ThrAlaArgTyrPheGlyIlePheAlaLeuI 156
895 GCCACCACACACATGATGCCGACCAACAGAGGGCGATGCAGATGCCGGT 846
156 eIleValGluThrLeuAlaTrpPheThrValAlaSerLeuPheAlaL 173
845 TATGGTCAGCACTCTCTCTGGTACAGCTCTCCGCTCCATAAATTCAA 796
173 euProGlnMetArg..... 177
795 TCCCAAGATGCTTGTAGAGCTGGCCATTACGTAGTTTGGCAGCGATCA 746
178 .....ArgGlyTyrGlnArgLeuAlaLysTyrPheAspGlyPhe 190
745 CCAGTAAACTCATTTGGGCACATTGCACAAGTATCTCGAGGGGTTT 701
seq_name: /cgn2_6/ptodata/2/lna/5B_COMB.seq:US-08-456-241-7
seq_documentation_block:
; Sequence 7, Application US/08456241
; Patent No. 5840525
; GENERAL INFORMATION:
; APPLICANT: Vandien, Richard L.
; APPLICANT: Holmes, William E.
; TITLE OF INVENTION: STRUCTURE, PRODUCTION AND USE OF
; TITLE OF INVENTION: HEREGULIN
; NUMBER OF SEQUENCES: 30
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Genentech, Inc.
; STREET: 460 Point San Bruno Blvd
; CITY: South San Francisco
; STATE: California
; COUNTRY: USA
; ZIP: 94080
; COMPUTER READABLE FORM:
; MEDIUM TYPE: 5.25 inch, 360 Kb floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: patin (Genentech)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/456,241
; FILING DATE: 31-MAY-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/126145
; FILING DATE: 23-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/880917
; FILING DATE: 11-MAY-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/847743
; FILING DATE: 06-MAR-1992
; PRIOR APPLICATION NUMBER: 07/790801
; FILING DATE: 08-NOV-1991
; PRIOR APPLICATION NUMBER: 07/765212
; FILING DATE: 25-SEP-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 07/705256
; FILING DATE: 24-MAY-1991
; ATTORNEY/AGENT INFORMATION:
;
; NAME: Lee, Wendy M.
; REGISTRATION NUMBER: 00,000
; REFERENCE/DOCKET NUMBER: 712P4C1D1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 415/225-1994
; TELEFAX: 415/952-9881
; TELEX: 910/371-7168
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 bases
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-456-241-7

alignment_scores:
Quality: 90.00 Length: 199
Ratio: 0.938 Gaps: 12
Percent Similarity: 48.241 Percent Identity: 26.131

alignment_block:
US-09-466-935-4 x US-08-456-241-7/rev ..
Align seg 1/1 to reverse of: US-08-456-241-7 from: 1 to: 2199
20 GlyProAspPhePheValSer...GlnThrAlaValSerArgSerAr 35
1202 GAAAAGGATGCTTTTCAGTGTGTCCTGCTCCAGCTGTGGCTAGGAGTCT 1153
35 GlysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMetVal 52
1152 GGGTCAGAGT...AGTGGAGTGTGGGC.....TGTGGAAAGTATAGTGA 1112
52 rPalaglyIleAlaLeuLeuGlyLeuHisLeu..... 62
1111 CTGGTGGAAAAGGATGTCTCTGCTCTCTCAACAATATGCTCAGTGA 1062
63 ...IleIleGluLysMetAlaTrpLeuHisLeuLeuIleMetValGly 78
1061 GATGACGTTTTAGATACGTATGTGATTCACACAGCTGGACATCTCGGGG 1012
78 yGlyLeuTyrLeuCysTrpMetGlyTyrGlnMetLeuArgGlyAlaLeu 95
1011 G.....TGGGTT.....AGGATGGTGAGG..... 983
95 yLysGluAlaValSerAlaProAlaProGlnValGluLeuAlaLysSer 111
992 .....CCCATTTGGCAATGTTTCATCATATTTGTTTCG 963
112 GlyArgSerPheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAl 128
962 TTCAGACCG.....AAGGCTCTGCCGAAGAC 937
128 alleIleTyrPheGlySerValPheSerLeuPheValGlyAspAsnValG 145
936 GGTCAAG.....CAGCTTTTCGCTGTTCTTCTGGTTTTCAGTAG 896
145 ly.Thr.....ThrAlaArgTyrPheGlyIlePheAlaLeuI 156
895 GCCACCACACACATGATGCCGACCAACAGAGGGCGATGCAGATGCCGGT 846
156 eIleValGluThrLeuAlaTrpPheThrValAlaSerLeuPheAlaL 173
845 TATGGTCAGCACTCTCTCTGGTACAGCTCTCCGCTCCATAAATTCAA 796
173 euProGlnMetArg..... 177
795 TCCCAAGATGCTTGTAGAGCTGGCCATTACGTAGTTTGGCAGCGATCA 746
178 .....ArgGlyTyrGlnArgLeuAlaLysTyrPheAspGlyPhe 190
745 CCAGTAAACTCATTTGGGCACATTGCACAAGTATCTCGAGGGGTTT 701
```

seq_name: /cgn2_6/ptodata/2/ina/5B_COMB.seq:US-08-550-815-7

seq_documentation_block:

```
; Sequence 7, Application US/08550815
; Patent No. 5869618
; GENERAL INFORMATION:
; APPLICANT: Lippman, Marc E
; APPLICANT: Lupu, Ruth
; TITLE OF INVENTION: Ligand Growth Factors that Bind to the
; TITLE OF INVENTION: erbB-2 Receptor Protein and Induce Cellular Response
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/550.815
; FILING DATE: 31-OCT-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,277
; FILING DATE: 26-JUL-1993
; APPLICATION NUMBER: US 07/875,788
; FILING DATE: 29-APR-1992
; APPLICATION NUMBER: US 07/640,497
; FILING DATE: 14-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,988
; FILING DATE: 24-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,114
; FILING DATE: 22-APR-1992
; APPLICATION NUMBER: US 07/528,438
; FILING DATE: 23-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoscheit, Dale H
; REGISTRATION NUMBER: 19,090
; REFERENCE/DOCKET NUMBER: 02899.43360
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-550-815-7
```

alignment_scores:
Quality: 90.00 Length: 199
Ratio: 0.938 Gaps: 12
Percent Similarity: 48.241 Percent Identity: 26.131

alignment_block:

US-09-466-935-4 x US-08-550-815-7/rev ..

Align seg 1/1 to reverse of: US-08-550-815-7 from: 1 to: 2199

```
20 GlyProAspPhePheValSer...GlnThrAlaValSerArgSerAr 35
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1202 GGAAGGATGCTTTCAGTGTGTCGCTGCTCCAGCTGGCTAGGACTCT 1153
35 gLySGluAlaMetMetGlyValLeuGlyLleThrCysGlyValMetValT 52
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1152 GGGTGACAGT...AGTGGAGTGATGGGC.....TGTGGAAGTATAGTGA 1112
52 rPalagLyIleAlaLeuLeuGlyLeuHisLeu..... 62
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1111 CTGGTGGAAAGGATGCTCTGCTTCTCTCAACAATATGCTCACTGGA 1062
63 ...llelleGluLysMetAlaTrpLeuHisThrLeuMetValGlyGl 78
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1061 GATGACGTTTTAGATACGATTGATTCACCAAGCTGGACATTCCTCGGGG 1012
78 yGlyLeuTyrrLeuCysTrpMetGlyTyrrGlnMetLeuArgGlyAlaLeuL 95
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1011 G.....TGGGT.....AGGATGGTCAGG..... 993
95 yLySGluAlaValSerAlaProAlaProGlnValGluLeuAlaLysSer 111
|||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
992 .....CCATTGGCAATGTTCAATCATATATTGTTTCG 963
112 GlyArgSerPheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAl 128
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
962 TTCAGACCG.....AAGGCTCTGCCGAAGAC 937
128 atlelleTyrrPheGlySerValPheSerLeuPheValGlyAspAsnValG 145
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
936 GGTCAATG.....CAGCTTTTCCGCTGTTCTTCGTTTTCAGTAG 896
145 ly.Thr.....ThrAlaArgTrpGlyLlePheAlaLeuL 156
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
895 GCCACCACACACATGATGCCGCCATTCAGTAGTTTGGCAGCGATCA 746
156 eileValGluThrLeuAlaTrpPheThrValValAlaSerLeuPheAlaL 173
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
845 TATGGTCAGCACTCTCTCTGTTAGTACAGCTCCTCGGCTCCATAAATCAA 796
173 euProGlnMetArg..... 177
795 TCCCAAGATGCTTGTAGAGCTGCCATTCAGTAGTTTGGCAGCGATCA 746
178 .....ArgGlyTyrrGlnArgLeuAlaLysTrpIleAspGlyPhe 190
||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
745 CCAGTAAACTCATTTGGCAGCTTGCACAAAGTATCTCGAGGGGTTT 701
seq_name: /cgn2_6/ptodata/2/ina/6A_COMB.seq:US-08-703-089-7
seq_documentation_block:
; Sequence 7, Application US/08703089
; Patent No. 6040290
; GENERAL INFORMATION:
; APPLICANT: Lippman, Marc E
; APPLICANT: Lupu, Ruth
; TITLE OF INVENTION: Ligand Growth Factors that Bind to the
; TITLE OF INVENTION: erbB-2 Receptor Protein and Induce Cellular Response
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Banner, Birch, McKie & Beckett
; STREET: 1001 G Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: U.S.
; ZIP: 20001
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/703.089
; FILING DATE: 31-OCT-1995
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/096,277
; FILING DATE: 26-JUL-1993
; APPLICATION NUMBER: US 07/875,788
; FILING DATE: 29-APR-1992
; APPLICATION NUMBER: US 07/640,497
; FILING DATE: 14-JAN-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/917,988
; FILING DATE: 24-JUL-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/872,114
; FILING DATE: 22-APR-1992
; APPLICATION NUMBER: US 07/528,438
; FILING DATE: 23-MAY-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoscheit, Dale H
; REGISTRATION NUMBER: 19,090
; REFERENCE/DOCKET NUMBER: 02899.43360
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-508-9100
; TELEFAX: 202-508-9299
; TELEX: 197430 BBMB UT
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2199 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: both
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; US-08-550-815-7
```



```

; COUNTRY: USA
; ZIP: 22202
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/311,023
; FILING DATE: 22-SEP-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: PATCH, Andrew J.
; REGISTRATION NUMBER: Reg. No. 5693465 32,925
; REFERENCE/DOCKET NUMBER: WCM.56
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 703/521-2297
; TELEFAX: 703/685-0573
; TELEX: 248425
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2404 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; US-08-311-023-3

alignment_scores:
    Quality: 86.50      Length: 179
    Ratio: 1.042      Gaps: 11
    Percent Similarity: 46.369      Percent Identity: 27.374

alignment_block:
US-09-466-935-4 x US-08-311-023-3 ..

Align seg 1/1 to: US-08-311-023-3 from: 1 to: 2404

53 AlaGlyIleAlaLeuLeuGlyLeuHisLeuIleGluLysMetAlaTr 69
|||||:||||| |||
787 GCCGGCTGCGCAGTTG.....GCCTG 809

69 pLeuHisThrLeuMetValGlyGlyLeuLeuLeuLeuLeuLeuLeu 86
|:|||||:||||| |||
810 G.....ATGCGTATAGGCGTATGCGTGCACAAATTCACCGATG 850

86 lyTyrGlnMetLeuArgGlyAlaLeuLysGlyAlaValSerAla... 101
||:|||||:||||| |||
851 GCCTAGCAAT.....GGTCTGCTTTTACTCAAGGCTTATCAAGTGGT 894

102 .....ProAlaProGln..... 105
|||||:

895 TTAAGTACTTCTGTGTTCTCTCATGAGTTGCGCTCATGAATTAGG 944
|||||:

106 .....ValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuL 120
||:|||||:||||| |||
945 TGACTTTGCTGTTCTACTAAGCGTGGCATGCCGTTAAGCGGCTGCC 994

120 euThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGlySerValPhe 136
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
995 TTTAATAAT...GCATTGTCAGCCATGCTGCGTATCTTGAATGCAACA 1041

137 SerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrpGlyIlePh 153
||:|||||:||||| |||
1042 GGAATTTTCATGTCATTATGCTGAATGTTCTATCTGG...ATATT 1088

153 eAlaLeuIleIleValGluThrLeuAlaTrpPheThrValValAlaSerL 170
|||||: ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
1089 TGCACCT.....ACTGCTGGCTATTATCATGATGTTGCTCTGG 1126

170 euPheAlaLeuProGlnMet..... 176
||:|||||:||||| |||
1127 TTGATAGGTACCTGAAATGCTGCACAATGATGCTAGTGACCATGGATGT 1176

```

```

177 ...ArgArgGlyTyrGlnArgLeuAlaLysTrpIleAspGlyPheAlaG1 192
||| ||||| ||| |||
1177 AGCGCGTGGGGTATTCTTTTACAGAAT.....GCTGG 1211

192 yAlaLeuPheAlaGlyPheGlyIleHisLeuIleIle 204
| |||: ||| ||||| ||| ||| |||
1212 GATGCTTTG...GGTTTGGAAATTATGTTACTATT 1245

seq_name: /cgn2_5/ptodata/2/lna/5A_COMB.seq:US-08-395-246C-1

seq_documentation_block:
; Sequence 1, Application US/08395246C
; Patent No. 5773214
; GENERAL INFORMATION:
; APPLICANT: Peery, Robert B.
; TITLE OF INVENTION: MULTIPLE DRUG RESISTANCE GENE OF
; TITLE OF INVENTION: ASPERGILLUS FLAVUS
; NUMBER OF SEQUENCES: 4
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Eli Lilly and Company
; STREET: Lilly Corporate Center
; CITY: Indianapolis
; STATE: IN
; COUNTRY: USA
; ZIP: 46285
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/395,246C
; FILING DATE:
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Plant, Thomas G.
; REGISTRATION NUMBER: 35784
; REFERENCE/DOCKET NUMBER: x9683
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 317-276-2459
; TELEFAX: 317-277-1917
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3924 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: cdna
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3924
; US-08-395-246C-1

alignment_scores:
    Quality: 85.50      Length: 173
    Ratio: 1.082      Gaps: 8
    Percent Similarity: 45.665      Percent Identity: 23.699

alignment_block:
US-09-466-935-4 x US-08-395-246C-1 ..

Align seg 1/1 to: US-08-395-246C-1 from: 1 to: 3924

3 MetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerPr 19
|||||:||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2347 CTCATGTTCTCATGATTGGCATTTGTCGCTGTGTGTATAGTGCGCA 2396

19 oglyProAspPhePhePheValSerGlnThrAlaValSerArgSerArgL 36
||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
2397 AGGACCTTGTGTGCATATAGTTCGAAAAGATGCTCTACCGGCGCGCA 2446

```

```
36 ysgluAa..... 38
2447 GTCAAGCATTCGGGTGATCTGACACAGGATATCTTTCTTCGATCAA 2496
38 ..... 38
2497 CAGGAAACACGACAGCGGCATTAAACAGCCACTCTGAGCGCGGGACCAA 2546
39 ....MetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAlaAg 54
2547 AGAACTAACTGACATTAGTGGCGTCACATTGGGACTATCTTGATTGTCT 2596
54 ly.....IleAlaLeuLeuGlyLeuHisLeuIleLeuGlyLeuMet 67
2597 CTGTCAACTCTGTGCTCTTCCTGGGGTACCCCTTGTAATA..... 2637
68 AlaTrpLeuHisThrLeuIleMetValGlyGly.....GlyLeuTyrLe 82
2638 GSCTGGAAGTTGGCACTGGTGTGATCTCCCGGTTTCCAGCCCTGCTGAT 2687
82 uCys.....TrpMetGlyTyrGlnMetLeuArgGlyAlaAl 94
2688 GTGTGATTTGTTCGCTTGGATTTGGAGCGGTTCCACGAGGCGCA 2737
94 eulys.....LysGluAlaValSerAla.ProAlaTrp 104
2738 AGAAACCATATCAAGAAATCAGCTAGTTCTGCTGTGAGGAGCCTCTGCC 2787
104 oGlnValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuLeuT 121
2788 ATCCGTA.....CGGTGCTTTCTTTGACCATGCGAGAC 2819
121 hrAsnLeuAlaAsnPro 126
2820 GGAGGCTTTGCAATCCT 2836
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-09-335-409-1
seq_documentation_block:
; Sequence 1, Application US/09335409
; Patent No. 6121029
; GENERAL INFORMATION:
; APPLICANT: Schupp, Thomas
; APPLICANT: Ligon, James
; APPLICANT: Molnar, Istvan
; APPLICANT: Zirkle, Ross
; APPLICANT: Cyr, Devon
; APPLICANT: Goerlach, Joern
; TITLE OF INVENTION: GENES FOR THE BIOSYNTHESIS OF EPOTHILONES
; FILE REFERENCE: 4-30582A
; CURRENT APPLICATION NUMBER: US/09/335,409
; CURRENT FILING DATE: 1999-06-17
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 1
; LENGTH: 68750
; TYPE: DNA
; ORGANISM: Sorangium cellulosum
US-09-335-409-1
```

```
alignment_scores:
  Quality: 82.00      Length: 209
  Ratio: 0.845      Gaps: 9
Percent Similarity: 46.411 Percent Identity: 22.010
alignment_block:
US-09-466-935-4 x US-09-335-409-1 ..
Align seq 1/1 to: US-09-335-409-1 from: 1 to: 68750
1 MetLeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMe 17
```

```
3439 GTCCCTGCTCTCGTCGACACCTCGCGCTCATCTCCGTCGTCGACGCGCGC... 3486
17 tSerProGlyProAspPhePheValSerGlnThrAlaValSerArgS 34
3487 .....GCCTCCGCGAGGCTCGCGCGCGGC 3511
34 erArgLysGluAlaMetMetGlyValLeu.....GlyIleThrCysGly 48
3512 TCGCGCAGCCCGAGGTCTCGGGAGCTCTCGCGCGGCTGCTGCTGGC 3561
49 ValMetValTrpAlaGlyIleAlaLeuLeuGlyLeuHis..... 61
3562 CCCTCCGCTC...GTGCGCGGCTCGCGCGCGGTTCCATCGAGCCCTCTT 3608
62 .....LeuIleIleGluLysMetAlaTrpLeuHisT 72
3609 CCAGAGCGCGCGGTCTCGGGTCTGCTCTCGGGCATCTCCTGG..... 3651
72 hrLeuIleMetValGlyGlyGlyLeuTyrLeuCysTrpMetGlyTyrGln 88
3652 .....ATAGCGCGCTCTCTCTGCTGCTGATGCGCGCATCGAG 3690
89 MetLeuArgGlyAlaLeuLysLysGluAlaValSerAlaProAlaProGl 105
3691 GTGAGCTGGGCATCTCTGCGCAAGGAGCG..... 3720
105 nValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLeuLeuThra 122
3721 .....CGCCCGCGGCGGCTCTCGCGCTCGGCGC 3748
122 snLeuAlaAsnProLysAlaIleIleTyrPheGlySerValPheSerLeu 138
3749 CGATCGCGCCCGCTCTCGCGCG.....GGCGCGCTCTCTCGCGC 3789
139 PheValGlyAspAsnValGlyThrThrAlaArgTrpGlyIlePheAlaLe 155
3790 CTCGCTCGCATCGGCCCTCTCGAGC.....GGCCTCTCTCTCGG 3830
155 rIleIleValGluThrLeuAlaTrpPheThrValValAlaSerLeuPheA 172
3831 GATCGTCTCTCGGTGACGCGCGGTGACGTATCGCGAGGTGCTGATCG 3880
172 laLeuProGlnMetArgGlyTyr..... 180
3881 AGCGGAGTCTGATCGCGCGCATATCGCGAGGTGACGCTCGCGCGCGG 3930
181 .....GlnArgLeuAlaLysTrpIle 187
3931 GTGGTACGAGGTCGCTGCTGGTG 3957
seq_name: /cgn2_6/ptodata/2/ina/6B_COMB.seq:US-08-940-424-10
seq_documentation_block:
; Sequence 10, Application US/08940424A
; Patent No. 6171837
; GENERAL INFORMATION:
; APPLICANT: Blauer, William S.
; APPLICANT: Zott, Roseann P.
; APPLICANT: Gamble, Mary V.
; APPLICANT: Mertz, James R.
; TITLE OF INVENTION: POTENT
; FILE REFERENCE: 0575/54544
; CURRENT APPLICATION NUMBER: US/08/940,424A
; CURRENT FILING DATE: 1997-09-29
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 10
; LENGTH: 1229
; TYPE: DNA
; ORGANISM: mouse
US-08-940-424-10
```

INHIBITORS OF HUMAN 9-cis RETINOL DEHYDROGENASE

```

alignment_scores:
  Quality: 79.00      Length: 209
  Ratio: 0.898       Gaps: 9
  Percent Similarity: 42.105  Percent Identity: 22.488

alignment_block:
US-09-466-935-4 x US-08-940-424-10  ..

Align seg 1/1  to: US-08-940-424-10  from: 1  to: 1229

12  HisIleValAlaLeuMetSerProGlyProAspPhePheValSerG1 28
|||||
93  CACATTGTGTG.....CTGCCAGCTTCCCCAGAGCCT 127

28  nThrAlaValSerArgLysGluAlaMetMetGlyValLeuGlyI 45
:::|||||:::  |||
128 AGCTGCCCTCAGCAGGCATCTCATCCCATCATGTGG.....CTGCCCTC 171

45  leThrCysGlyValMetValTPAlaGlyIleAlaLeuLeuGlyLeuHis 61
:::|||||:::|||||:::|||||:::
172 TGCTTCGGGTGCTTGCCTGTGGCGAGTGTGTGTGCTCAGAGACCGG 221

62  LeuIleIleGluLysMetAlaTrpLeuHisThrLeuIleMetValGlyI 78
:::|||||:::|||||:::|||||:::
222 .....CAGAGCCTGCCGCCAGTGTGTGTTCATCTTCATCACTGG 262

78  Y.....GlyLeuTyrLeuCystTrpMetG 86
|:::|||||
263 CTGTGACTCTGGCTTTGGGCGCTTCTGGCACTGCACTTGACCCAGAAGG 312

86  lyTyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSerAlaPro 102
|||||:::|||||:::|||||:::
313 GCITCCAGTCCGCGCGCTGCTG.....ACCCCC 344

103 AlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyLe 119
:::|||||:::|||||:::
345 TCTGGAGCAGAGACCTGCGAGCATGGCTCTCCCGCCTCCACACAAC 394

119 uLeuThrAsnLeuAlaAsnProLysAlaIleIleTyrPheGlySerValP 136
{|||:::|||||:::
395 ACTACTGGATATCATCATCCCCAG..... 419

136 heSerLeuPheValGlyAspAsnValGlyThrThrAlaArgTrp..... 150
|||||:::|||||:::
420 .....ATGTCACCAAGTTGCCAAGTGGGTGAAG 449

151 .....GlyIlePheAlaLeuIle..... 156
|||||:::|||||:::
450 ACAGCTTTGGAGAACTGGACTTTTGTGCTGGTGAATAACGCTGGCGT 499

157 .....IleValIleuThrLeuAlaTrpPheThr..... 165
|||||:::|||||:::
500 AGCTGGTATCATCGGCCCCACACCATGGCTGAACACAGGATGATTCCAGA 549

166 .....ValValAlaSerLeu 170
|||||:::|||||
550 GAGTACTGAGTGTGAACACTGGGGCCCATCGGTGTACCCCTTGGCCCTG 599

171 PheAlaLeuProGlnMetArgGly 179
:::|||||
600 CTGCCCTGTACAGCAGGCCAGGGGT 626

```

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:40:01 ; Search time 26.05 Seconds
(without alignments)
543,452 Million cell updates/sec

Title: US-09-466-935-4
Perfect score: 1054
Sequence: 1 MLMFLTVAMVHIVALMSGP.....IDGFAGALFAGFGIHLIISR 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 19801 seqs, 68722935 residues

Total number of hits satisfying chosen parameters: 198801

Minimum DB seq length: 0
Maximum DB seq length: 2000000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_67: *
1: pir1: *
2: pir2: *
3: pir3: *
4: pir4: *

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	618	58.6	122	2 H65186	hypothetical 13.3
2	368.5	35.0	210	2 F84115	hypothetical prote
3	313.5	29.7	222	2 F82353	conserved hypothet
4	292	27.7	210	2 D84016	hypothetical prote
5	274.5	26.0	204	2 B83279	hypothetical prote
6	269.5	25.6	212	2 T43921	yfud protein [impo
7	263.5	25.0	213	2 F83444	hypothetical prote
8	261.5	24.8	223	2 H64759	membrane protein y
9	246.5	23.4	204	2 F83306	hypothetical prote
10	245	23.2	210	2 G83082	hypothetical prote
11	208.5	19.8	206	2 G82979	hypothetical prote
12	207	19.6	216	2 F83051	conserved hypothet
13	201	19.1	207	2 B83703	hypothetical prote
14	199	18.9	216	2 C55580	hypothetical prote
15	197.5	18.7	204	2 B82410	conserved hypothet
16	190	18.0	212	2 F84940	hypothetical prote
17	186	17.6	208	2 G84086	hypothetical prote
18	180.5	17.1	210	2 F69975	dihydrodipicolinat
19	174.5	16.6	241	2 C75329	conserved hypothet
20	172	16.3	212	2 G82200	conserved hypothet
21	167.5	15.9	209	2 E82388	conserved hypothet
22	163	15.5	213	2 C82523	amino acid transpo
23	159	15.1	203	2 T30317	chemotaxis protein
24	159	15.1	206	2 S76178	hypothetical prote
25	155.5	14.8	208	2 C82471	conserved hypothet
26	142.5	13.5	205	2 G82358	conserved hypothet
27	141.5	13.4	206	2 C82139	conserved hypothet
28	136	12.9	207	2 D83187	hypothetical prote
29	123.5	11.7	208	1 B69066	conserved hypothet

ALIGNMENTS

RESULT 1

H65186
hypothetical 13.3 kD protein in recQ 3' region - Escherichia coli (strain K-12)
N:Alternate names: hypothetical protein ol28
C:Species: Escherichia coli
C:Date: 12-Sep-1997 #sequence_revision 17-Sep-1997 #text_change 08-Oct-1999
C:Accession: H65186; S30713
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.,
.A.; Rose, D.J.; Mau, B.; Shao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617
A:Accession: H65186
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-122 <BLAT>
A:Cross-references: GB:AE000458; GB:U00096; NID:q2367299; PIDN:AAC76826.1; PID:q23673
A:Experimental source: strain K-12, substrain MG1655
R:Daniels, D.L.; Plunkett III, G.; Burland, V.; Blattner, F.R.
Science 257, 771-778, 1992
A:Title: Analysis of the Escherichia coli genome: DNA sequence of the region from 84.
A:Reference number: S30660; MUID:92358234
A:Accession: S30713
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 'YAVSLLD',2,'LPDATWCTE',12,'RGGFCTCATGRAGEKWAQ',31,'PES',35-122 <DAN>
A:Cross-references: EMBL:M87049
A:Note: this sequence has been corrected
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, November 1992
C:Genetics:
A:Gene: ylgJ

Query Match 58.6%; Score 618; DB 2; Length 122;
Best Local Similarity 100.0%; Pred. No. 2.6e-48;
Matches 122; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 85 MGYOMLRGALKKEAVSAPAPQVELAKSGRSFLKGLLTNLNPKAIYFGSVSLFVGDNV 144

Db 1 MGYOMLRGALKKEAVSAPAPQVELAKSGRSFLKGLLTNLNPKAIYFGSVSLFVGDNV 60

Qy 145 GTTARNGIFALIIIVETLAWFTVVASLFPQMRGYYORLAKWIDGFAGALFAGFGIHLII 204

Db 61 GTTARNGIFALIIIVETLAWFTVVASLFPQMRGYYORLAKWIDGFAGALFAGFGIHLII 120

Qy 205 SR 206

Db 121 SR 122

RESULT 2

F64115

hypothetical protein H11307 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 29-Sep-1999
C:Accession: F64115
R:Fietschmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage, A.; Gossyne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, J.D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID: 95350630
A:Accession: F64115
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-210 <TIGR>
A:Cross-references: GB:U32810; GB:I42023; NID:g1574760; PIDN:AAC22954.1; PID:g1574766; MUID:20437337
C:Genetics:
A:Start codon: GTG
C:Superfamily: hypothetical protein bl798

Query Match 35.0%; Score 368.5; DB 2; Length 210;
Best Local Similarity 38.8%; Pred. No. 8.1e-26;
Matches 80; Conservative 42; Mismatches 73; Indels 11; Gaps 4;

QY 4 LFTVAMVHIVAMSPGDPFFVSGTAVSRKRAMGVLCITCGVMWAGIALGLHLI 63
Db 1 MMLNLIIVHFLGTPGDPFFVSRMAASNRRTVCGILGILTLGIAFWGMSLGLAVL 60
QY 64 IEKMAWLHTLIMVGGGLYLCWGYQMLRGA--LKKEAVSAPAPQVELAKS---GRSFLKGL 120
Db 61 FVTIPALHGVIMLGGSLAYLGLFMARS--KKYKPFESHSDTEFNQQTIKKILKGL 118
QY 121 TNLANPKAIYFGVSFLFVGDNVGTTARWGI---FALLIVETLAWFTVVASLFPALPQMR 177
Db 119 VNLUNAKVVYESSVNSLVL---VNITEMQOILLAFVIVVFYFVVISUIFSRNIAK 175
QY 178 RGYQRLAKWIDGFAGALFAGFGIHLI 203
Db 176 RLYSQYSRYIDNMAGIVFLFGCVLV 201

RESULT 3
F82353
conserved hypothetical protein VC0191 [imported] - Vibrio cholerae (group O1 strain N169)
C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: F82353
R:Reidberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.; Chardson, D.; Ermolaeva, M.B.; Vamathevan, J.; Bass, S.; Qin, H.; Dragol, I.; Sellers, P. L. R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M. Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: F82353
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-222 <HEI>
A:Cross-references: GB:AE004109; GB:AE003852; NID:g9654590; PIDN:AAF93367.1; GSPDB:GN001
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0191
A:Map position: 1

Query Match 29.7%; Score 313.5; DB 2; Length 222;
Best Local Similarity 31.3%; Pred. No. 6.9e-21;
Matches 62; Conservative 52; Mismatches 79; Indels 5; Gaps 2;

QY 2 LMLFTVAMVHIVAMSPGDPFFVSGTAVSRKRAMGVLCITCGVMWAGIALGLH 61
Db 4 MSILATLAGVHFIALSPGDPVALVQNAHQGRKTKGVMTLGLSGILVHLIISLGS 63

QY 62 LIEKMAWLHTLIMVGGGLYLCWGYQMLRGA--LKKEAVSAPAPQVELAKSGRSFLK 117
Db 64 YLVVQOPMFLNLQLAGSSVLLYGALQSVMAQKNASTPHTSPAPSI-LGNRRQAFTK 122
QY 118 GLLTNLANPKAIYFGVSFLFVGDNVGTTARWGI--FALLIVETLAWFTVVASLFPALPQMR 177
Db 123 GWMNLLNPKALVFFVSLSSLLIPASMSVSGKVSAAAILVGLSLTWFSCLAWLLTTSAMQ 182
QY 178 RGYQRLAKWIDGFAGALF 195
Db 183 QRMQRITRSVDSICAAVF 200

RESULT 4
D84016
hypothetical protein BH2932 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: D84016
R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fuji, F.; H. Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans a
A:Reference number: A83650; MUID:20263314
A:Accession: D84016
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AP001517; GB:BA000004; NID:g10175500; PIDN:BAB06651.1; GSPDB:G
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH2932
C:Superfamily: hypothetical protein bl798

Query Match 27.7%; Score 292; DB 2; Length 210;
Best Local Similarity 32.4%; Pred. No. 5.4e-19;
Matches 67; Conservative 37; Mismatches 97; Indels 6; Gaps 2;

QY 4 LFTVAMVHIVAMSPGDPFFVSGTAVSRKRAMGVLCITCGVMWAGIALGLHLI 63
Db 1 MFVEFVIVGLLAGSPGDPFFIVNKNLSLFGARVGLTSLGASALIVHIVITVVLGFAFL 60
QY 64 IEKMAWLHTLIMVGGGLYLCWGYQMLRGA--LKKEAVSAPAPQVELAKSGRS---FLK 117
Db 61 IETYPALFFTIQLLAGAAYLWLGFHAIRSSPPKKEAEIEETQPIQSTKDSKSSIQGFK 120
QY 118 GLLTNLANPKAIYFGVSFLFVGDNVGTTARWGI--FALLIVETLAWFTVVASLFPALPQMR 177
Db 121 GFITNLLNPKALFFLSIFSQFTPQTADWVRWYGLVWVAVGLWFSFLAIFISYKHF 180
QY 178 RGYQRLAKWIDGFAGALFAGFGIHLII 204
Db 181 RPYQTHSYWFDRELGAALLFFAIRIII 207

RESULT 5
B83279
hypothetical protein PA2929 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: B83279
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L. ; Lory, S.; Olson, M.V. Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: B83279
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <STO>
A:Cross-references: GB:AE004719; GB:AE004091; NID:g9949021; PIDN:AAG06317.1; GSPDB:GN
A:Experimental source: strain PA01


```
Db 20 VYLTVGL-FVITTFNPGANLFFVVVQTSLSARRAGVLTGLGVALGDADFVSGIGLGLATL 78
Qy 64 IEKMWLHTLIMVGGGLYLCWMGYOMLRGALKREAVSAPAPQVELAKSGRS-----FLK 117
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 79 ITOCEEISLIRIVGAILWFAFNCMR-----RQSTQMSITLOQPIAPWVVFRR 130
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 118 GLITNLNAPKAIYFGSVFSLFVGDNVGTTARWIGIFALIIVETLAWFTVVASLFAIPOMR 177
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 131 GLITLSNQTVLFFISFVTLNAPETPTWRLMAMAGIVLASIIRVFLSQAFSLPAVR 190
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 178 RGYORLAKWIDGFAGALFAGFGIHLI 203
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 191 RAYGRMORVASRVIGIIGVFALRLI 216
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 9
F83306
hypothetical protein PA2710 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83306
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
A:Reference number: A82950; MUID:20437337
A:Accession: F83306
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <STO>
A:Cross-references: GB:AE004699; GB:AE004091; NID:99948782; PIDN:AAG06098.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA2710
C:Superfamily: hypothetical protein bl798

Query Match 23.4%; Score 246.5; DB 2; Length 204;
Best Local Similarity 23.2%; Pred. No. 6e-15;
Matches 62; Conservative 35; Mismatches 100; Indels 15; Gaps 2;

Qy 1 MLMFLTAMVHIVHIALMSFGPDFEFVSQTAVSRSRKEAMMGVLGITCGVMVWAGIALLGL 60
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1 MLTSLFVATLATLGMSPGPDFELIIRNAARYORSAAWMTSLGVILGVATHMAYCVAGL 60
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 61 HLTIERMAWLHTLIMVGGGLYLCWMGYOMLRGALKREAVSAPAPQVELAKSG-----R 113
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 61 AVLITTPWLFNALXYTGAVLIWIGIQALR-----SRGGTLDLAVGVQVRVGHWS 112
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 114 SFKGLITNLNAPKAIYFGSVFSLFVGDNVGTTARWIGIFALIIVETLAWFTVVASLFA 173
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 113 AFLOGYLCNLLNPRATLEFLAVFTQVLSLQSSFAEKLTAGIIVGLVIMWPLIWLIIQS 172
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 174 PQMRGQYORLAKWIDGFAGALFAGFGIHLIS 205
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 173 AVVRSRLARQAQGVVDKLLGGLLIALGVKVALS 204
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 10
G83082
hypothetical protein PA4507 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: G83082
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho-
A:Reference number: A82950; MUID:20437337
A:Accession: G83082
A:Status: preliminary
```

```
A:Molecule type: DNA
A:Residues: 1-210 <STO>
A:Cross-references: GB:AE004864; GB:AE004091; NID:99950740; PIDN:AAG07895.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4507
C:Superfamily: hypothetical protein bl798

Query Match 23.2%; Score 245; DB 2; Length 210;
Best Local Similarity 33.3%; Pred. No. 8.4e-15;
Matches 75; Conservative 34; Mismatches 68; Indels 48; Gaps 12;

Qy 6 LTVAMVHIVHIALMSFGPDFEFVSQTAVSRSRKEAMMGVLGITCGVMVWAGIALLGLH 61
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 8 IALAVYLV-----PGDMLLLFOTGARQGRRAALVTALGLARACHVL-----MAATGLA 59
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 62 LIITERMAWLHTLIMVGGGLYLCWMGYOMLRG---AL-KKEAVSAP-APQVELAKSGRSFL 116
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 60 LLFRTAPWTFDLVRLIGAVYLAWLGLQMLRGGLALFTSDAGSAPVPHAD-----RRALL 115
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 117 KGLITNLNAPKAIYFGSVFSLFVGDNVGTTARWIGIFALIIVETLAWFTVVASL----- 170
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 116 RGLITNLNAPKALFCSVLLPQFVSPAGSLA-----VQFAALGTVLVLVCLAFDC 166
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 171 -PALPQMRG-----YORLAKWIDGFAGALFAGFGIHLIISR 206
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 167 AYALAGRLGRWLASRPRAQLQW--GF-GLLIGFVRLALLR 208
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |

RESULT 11
C82979
hypothetical protein PA5341 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: C82979
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Bro-
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; L
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pa
A:Reference number: A82950; MUID:20437337
A:Accession: C82979
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-206 <STO>
A:Cross-references: GB:AE004946; GB:AE004091; NID:99951650; PIDN:AAG08726.1; GSPDB:GN
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA5341
C:Superfamily: hypothetical protein bl798

Query Match 19.8%; Score 208.5; DB 2; Length 206;
Best Local Similarity 28.8%; Pred. No. 1.5e-11;
Matches 65; Conservative 32; Mismatches 86; Indels 43; Gaps 9;

Qy 1 MLMFLTAMVHIVHIALMSFGPDFEFVSQTAVSRSRKEAMMGVLGITCGVMVWAGIALLGL 60
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1 MEAFVLYVASTHFAALLSPGPDFELVRAALLRRRQAD---GVAAGI---ALANLLSM 53
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 61 HLTIERMA-----WLHTLIMVGGGLYLCWMGYOMLRGALKREAVSAPAPQVELA 109
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 54 LVLVGLLASVPDSAHSAURLQAL---CGLYFLWLAGOAL---LAQRLEMPAQRDVPS 106
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 110 KSG--RSFLKGLITNLNAPKAIYFGSVFSLFVGDNVGTTAR-----WGI---FALIVE 159
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 107 RGYLRGLRDLGLASSLNPKLPFYAGLF-----GVLARFSLPGNALCLAWNSLA 158
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Qy 160 TLAWFTVVASLFAIPOMRQYORLAKWIDGFAGALFAGFGIHLIS 205
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
Db 159 VLGWDVALIVRLDRFRWGRVQRRVGVGALDRLCGVLLALLGGWLVIA 204
| : : : : : | : : : : : | : : : : : | : : : : : | : : : : : |
```



```

RESULT 12
F83051
conserved hypothetical protein PA4757 [imported] - Pseudomonas aeruginosa (strain PA01)
C:Species: Pseudomonas aeruginosa
C>Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000
C:Accession: F83051
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warren, P.; Hickey, M.J.; B.
adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim,
.; Lory, S.; Olson, M.V.
Nature 406, 959-964, 2000
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic patho
A:Reference number: A82950; MUID:20437337
A:Accession: F83051
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <STO>
A:Cross-references: GB:AE004889; GB:AE004091; NID:g9951014; PIDN:AAG08143.1; GSPDB:GN001
A:Experimental source: strain PA01
C:Genetics:
A:Gene: PA4757
C:Superfamily: hypothetical protein b1798

Query Match 19.6%; Score 207; DB 2; Length 216;
Best Local Similarity 29.4%; Pred. No. 2.1e-11;
Matches 64; Conservative 30; Mismatches 88; Indels 36; Gaps 4;

QY 5 FLTVAMVHVIALMSGPDPFFVVSQTAVSRKKEAMGVLTGCGVWVWAGIALGLHLII 64
DB 9 FWTYVLGVVVFVILLPGPNSLFLVATSAQRGVATGYRAACGVFLGDAVLMLLSALGVASLL 68

QY 65 EKMAWLHTLMVGGGLYLCWGYOMLRGALKK----EAVSAPAPQVELAKSGRSFLKGL 120
DB 69 KAEPMLFGLKYLGAAYLFYLGVMGLRGAWKRLNPEATAGAAEQVDV---HQPFRKALL 125

QY 121 TNLANKPAIIYFGSVFSLFVGDVNTTARMGIFALIIVETL-----AWF 164
DB 126 LSLSNPKALFFISFFIQFVDPGYAYPGLSFLVLAIVLELVLSALYLSFLIFTGVRLAAWF 185

QY 165 TVVASLFPALPQMRGYORLAKWIDGFAGALFAGFGIHL 202
DB 186 -----RRQRLLAAGATSGVGFVGVK 210

RESULT 13
E83703
hypothetical protein BH0429 [imported] - Bacillus halodurans (strain C-125)
C:Species: Bacillus halodurans
C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 31-Dec-2000
C:Accession: E83703
R:Takami, R.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira
Nucleic Acids Res. 28, 4317-4331, 2000
A:Title: Complete genome sequence of the alkaliphilic bacterium Bacillus halodurans and
A:Reference number: A83650; MUID:20263314
A:Accession: E83703
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-207 <STO>
A:Cross-references: GB:AP001508; GB:BA000004; NID:g10172890; PIDN:BAH04148.1; GSPDB:GN00
A:Experimental source: strain C-125
C:Genetics:
A:Gene: BH0429

Query Match 19.1%; Score 201; DB 2; Length 207;
Best Local Similarity 28.3%; Pred. No. 7e-11;
Matches 63; Conservative 39; Mismatches 81; Indels 40; Gaps 9;

QY 2 LMLFLTVAMVHVIALMSGPDPFFVVSQTAVSRKKEAMGVLTGCGVWVWAGIALGLHL 61
DB 6 LLSFLGVA---VLLTLMGPDILFVLAQSMNQFAGIVTALGLCTGLLVHITAAVGVIS 62

```

```

QY 62 LIIEKMAWLHTLMVGGG---LYLCWGYOMLRGALKKEAVSAPAPQVELAKSGRSFLK 118
DB 63 AIUYOSALAFVTVKAGAAVLLYLAWKAPQE-----KGEGLSID-KOTTIA-YGALYKKG 115

QY 119 LUTNLANKPAIIYFGSVFSLFVGDVNTTARMG-----IFALIIVETLAWFT---- 165
DB 116 IIMNVLNPKVSLFFLALLPOFVNSGAG-SAPQMQLLGVVFLIQAFIIFSLVSWFAEKVG 174

QY 166 --VVASLFPALPQMRGYORLAKWIDGFAGALFAGFGIHLISR 206
DB 175 QLLMRSSPIKQMR-----IRGGLLALIGLQVAFSK 206

RESULT 14
C55580
hypothetical protein (carA 3' region) - Pseudomonas aeruginosa
C:Species: Pseudomonas aeruginosa
C>Date: 25-Aug-1995 #sequence_revision 25-Aug-1995 #text_change 29-Sep-1999
C:Accession: C55580
R:Kwon, D.H.; Lu, C.D.; Walthall, D.A.; Brown, T.M.; Houghton, J.E.; Abdelal, A.T.
J. Bacteriol. 176, 2532-2542, 1994
A:Title: Structure and regulation of the carAB operon in Pseudomonas aeruginosa and P
A:Reference number: A55580; MUID:94222830
A:Accession: C55580
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-216 <KWO>
A:Cross-references: GB:004992; NID:g451649; PIDN:AAA19047.1; PID:g451652
C:Superfamily: hypothetical protein b1798

```

```

Query Match 18.9%; Score 199; DB 2; Length 216;
Best Local Similarity 28.4%; Pred. No. 1.1e-10;
Matches 62; Conservative 31; Mismatches 89; Indels 36; Gaps 4;

QY 5 FLTVAMVHVIALMSGPDPFFVVSQTAVSRKKEAMGVLTGCGVWVWAGIALGLHLII 64
DB 9 FWTYVLGVVVFVILLPGPNSLFLVATSAQRGVATGYRAACGVFLGDAVLMLLSALGVASLL 68

QY 65 EKMAWLHTLMVGGGLYLCWGYOMLRGALKK----EAVSAPAPQVELAKSGRSFLKGL 120
DB 69 KAEPMLFGLKYLGAAYLFYLGVMGLRGAWKRLNPEATAGAAEQVDV---HQPFRQALL 125

QY 121 TNLANKPAIIYFGSVFSLFVGDVNTTARMGIFALIIVETL-----AWF 164
DB 126 LSLSNPKALFFISFFIQFVDPGYAYPGLSFLVLAIVLELVLSALYLSFLIFTGVRLAAWF 185

QY 165 TVVASLFPALPQMRGYORLAKWIDGFAGALFAGFGIHL 202
DB 186 -----RRQRLLAAGATSGVGFVGVK 210

```

```

RESULT 15
B82410
conserved hypothetical protein VCA0846 [imported] - Vibrio cholerae (group O1 strain
C:Species: Vibrio cholerae
C>Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 15-Sep-2000
C:Accession: B82410
R:Heidelberger, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A:Reference number: A82035; MUID:20406833
A:Accession: B82410
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-204 <HEI>
A:Cross-references: GB:AE004412; GB:AE003853; NID:g9658269; PIDN:AAF96744.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VCA0846
A:Map position: 2

```

```
Query Match      18.7%; Score 197.5; DB 2; Length 204;
Best Local Similarity 25.2%; Pred. No. 1.4e-10;
Matches 51; Conservative 46; Mismatches 102; Indels 3; Gaps 1;

QY 2 LMLRLTYAMVHIYALKSPGDFEFYSOTAVSRKRKEANMMGYLGITCGVMYWGIALIGLH 61
Db 3 LTVVLSLFTICILGAMSPGSLANWAKHSLAGGRANGFAAAWAHAFGIGVYAFITLIGLA 62
QY 62 LIIEKMAWLHTLIMVGGGLYLCWMGYQMLRGALKKEAVSAPAPQVELAKSGRSLKGLLT 121
Db 63 VVLHQSPVLFRTISYAGAAYLAYLGNALR---SKGGVAAKLESGESVSWQSAAREGLLI 119
QY 122 NLAMPKAIYFGSVFSLFVGDNVGTARWGIFALIIVETLAWFTTWVASLFPALPOMRRGYQ 181
Db 120 SLLSPKIALFFIALFSQYVAVGSDLTSKAAIVITPLVVDGLWYSFITLILSPRLLOKLR 179
QY 182 RLAKWIDGFAGALFAGFGIHLI 203
Db 180 ARAVLIDRLSGVLIALAIRVL 201
```

Search completed: May 6, 2001, 14:40:02
Job time: 343 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: May 6, 2001, 14:41:20 ; Search time 16.79 Seconds
(without alignments)
420.288 Million cell updates/sec

Title: US-09-466-935-4

Perfect score: 1054

Sequence: 1 MLMLFTVAMVHVALMSPG.....IDGFAGALFAGFGIHLIISR 206

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 93435 seqs, 34255486 residues

Total number of hits satisfying chosen parameters: 93435

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Swissprot_39.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	DB ID	Description
1	1054	100.0	206	1 RHTC-ECOLI	P27846 escherichia
2	368.5	35.0	210	1 YD07-HAEBIN	Q57320 haemophilus
3	261.5	24.8	223	1 YAHN-ECOLI	P75693 escherichia
4	207	19.6	216	1 YBF7-PSEAE	P38102 pseudomonas
5	190	18.0	212	1 YEAS-ECOLI	P76129 escherichia
6	180.5	17.1	210	1 YH2P-BACSU	O05406 bacillus su
7	159	15.1	203	1 CHPE-PSEAE	O87005 pseudomonas
8	144.5	13.7	206	1 YG27-STNY3	P74343 synechocyst
9	141.5	13.7	206	1 RHTB-ECOLI	P27847 escherichia
10	111.5	10.6	206	1 YGGA-AERSA	P70775 aeromonas s
11	105	10.0	236	1 LYSE-CORGL	P94633 corynebacte
12	104.5	9.9	225	1 YGGA-AERHY	P52047 aeromonas h
13	103	9.8	211	1 YGGA-ECOLI	P1667 escherichia
14	101	9.6	195	1 YF1K-ECOLI	P38101 escherichia
15	96.5	9.2	201	1 Y488-MYCTU	Q11154 mycobacteri
16	92	8.7	405	1 WCAD-ECOLI	P71238 escherichia
17	89	8.4	199	1 YJ86-MYCTU	Q10871 mycobacteri
18	88	8.3	541	1 COX1-BRAJA	P31833 bradyrhizob
19	87.5	8.3	722	1 VAT1-HALN1	Q9hnd8 halobacteri
20	85.5	8.1	695	1 DIP2-MYCTU	Q10801 mycobacteri
21	84.5	8.0	459	1 NU4M-POLAR	Q95917 polypterus
22	84.5	8.0	663	1 CYOB-ECOLI	P18401 escherichia
23	83.5	7.9	193	1 ISP2-VITSL	O9xds0 vitreoscill
24	83	7.9	464	1 GNP1-ECOLI	P27021 caenorhabdi
25	82.5	7.8	653	1 YTH3-CARBL	Q26337 methanobact
26	82	7.8	216	1 RIB7-METTH	P34196 crossstoma
27	81	7.7	173	1 NU6M-CHROLA	Q49131 methylobact
28	80.5	7.6	289	1 MAUN-METEX	P31135 escherichia
29	80.5	7.6	299	1 YSW-ECOLI	P31125 escherichia
30	80.5	7.6	299	1 EAMA-ECOLI	P31125 escherichia
31	80.5	7.6	299	1 YDED-ECOLI	P31125 escherichia
32	80.5	7.6	1080	1 CYAT-HUMAN	P51828 homo sapien
33	80	7.6	607	1 NUSM-CARAU	O78688 carassius a

34	79.5	7.5	379	1	CYB-AKOBO	P21714 akodon boli
35	79.5	7.5	379	1	CYB-DASNO	O21337 dasypus nov
36	79.5	7.5	459	1	NU4M-HALGR	P38601 halichoerus
37	79	7.5	172	1	NU6M-CYPECA	P24982 cyprinus ca
38	79	7.5	452	1	NU4M-BRAFL	O47423 branchiost
39	79	7.5	452	1	NU4M-BRALA	O79421 branchiost
40	78.5	7.4	133	1	PTPW-ECOLI	P42905 escherichia
41	78.5	7.4	173	1	NU6M-CARAU	O78689 carassius a
42	78.5	7.4	662	1	CYOB-BOCAI	P57543 buchnera ap
43	78	7.4	493	1	SECY-ARCFU	O28377 archaeoglob
44	78	7.4	534	1	COX1-KUGLA	P20386 kluyveromyc
45	77.5	7.4	262	1	COX3-METSE	O47491 metridium s

ALIGNMENTS

RESULT	1
RHTC-ECOLI	
ID	RHTC-ECOLI
AC	P27846
DT	01-AUG-1992 (Rel. 23, Created)
DT	15-JUL-1998 (Rel. 36, Last sequence update)
DT	15-JUL-1999 (Rel. 38, Last annotation update)
DE	THREONINE EFFLUX PROTEIN.
GN	RHTC.
OS	Escherichia coli.
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC	Escherichia.
OX	NCBI_TaxID=562;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN-K12 / MG1655;
RX	MEDLINE=92358234; PubMed=1379743;
RA	Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT	"Analysis of the Escherichia coli genome: DNA sequence of the region
RT	from 84.5 to 86.5 minutes.";
RL	Science 257:771-778(1992).
RN	[2]
RP	REVISIONS.
RC	STRAIN-K12 / MG1655;
RX	MEDLINE=97426617; PubMed=9278503;
RA	Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA	Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA	Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA	Mau B., Shao Y.;
RT	"The complete genome sequence of Escherichia coli K-12.";
RL	Science 277:1453-1474(1997).
RN	[3]
RP	SEQUENCE OF 1-107 FROM N.A.
RC	STRAIN-K12;
RX	MEDLINE=87115164; PubMed=3027506;
RA	Irino N., Nakayama K., Nakayama H.;
RT	"The recQ gene of Escherichia coli K12: primary structure and
RT	evidence for SOS regulation.";
RL	Mol. Gen. Genet. 205:298-304(1986).
RN	[4]
RP	CONCEPTUAL TRANSLATION.
RA	Rudd K.E.;
RL	Unpublished observations (DEC-1997).
RN	[5]
RP	CHARACTERIZATION.
RX	MEDLINE=99313167; PubMed=10386596;
RA	Zakataeva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
RA	Livshits V.A.;
RT	"The novel transmembrane Escherichia coli proteins involved in the
RT	amino acid efflux.";
RL	FEBS Lett. 452:228-232(1999).
CC	-!- FUNCTION: CONDUCTS THE EFFLUX OF THREONINE.
CC	-!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC	-!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC	-!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT WAS
CC	INTRODUCED IN POSITION 80 TO PRODUCE THIS ORF.

<p>-!- CAUTION: REF_3 SEQUENCE DIFFERS FROM THAT SHOWN DUE TO FRAMESHIFTS IN POSITIONS 47 AND 73.</p>						
<p>-----</p>						
<p>This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL Outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/) or send an email to license@sb-sib.ch.</p>						
<p>-----</p>						
CC	EMBL;	M87049;	AAA67619.1;	ALT_FRAME.		
DR	EMBL;	AB000458;	AAC76826.1;	ALT_FRAME.		
DR	EMBL;	M30198;	-;	NOT_ANNOTATED_CDS.		
DR	PIR;	S30713;	S30713.			
DR	EcoGene;	EG11468;	rhtC.			
DR	InterPro;	IPR001123;	-;			
DR	Pfam;	PF01810;	LyseE; 1.			
KW	Transport;	Transmembrane.				
FT	TRANSMEM	1	21	POTENTIAL.		
FT	TRANSMEM	44	64	POTENTIAL.		
FT	TRANSMEM	67	87	POTENTIAL.		
FT	TRANSMEM	150	173	POTENTIAL.		
SEQ	SEQUENCE	206 AA;	22474 MW;	F64017878CC6D50D CRC64;		
<p>-----</p>						
<p>Query Match 100.0%; Score 1054; DB 1; Length 206;</p>						
<p>Best Local Similarity 100.0%; Pred No. 2.9e-77;</p>						
<p>Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;</p>						
<p>-----</p>						
QY	1	MLMFLTVARVHVIALMSGPDFPFSQTAVSRSRKEAMMGVLITCGVMWAGIALLGL	60			
DB	1	MLMFLTVARVHVIALMSGPDFPFSQTAVSRSRKEAMMGVLITCGVMWAGIALLGL	60			
<p>-----</p>						
QY	61	HLIIKMAWLHTLIWMGGYLICWNGYQMLRGALKKEAVSAPAPQVELAKSGRSFKGLL	120			
DB	61	HLIIKMAWLHTLIWMGGYLICWNGYQMLRGALKKEAVSAPAPQVELAKSGRSFKGLL	120			
<p>-----</p>						
QY	121	TNLANPKAIIYFGSVFSLFVGDNVGTTRAWGFALIIIVETLAWTTVVASLFALPMRRGY	180			
DB	121	TNLANPKAIIYFGSVFSLFVGDNVGTTRAWGFALIIIVETLAWTTVVASLFALPMRRGY	180			
<p>-----</p>						
QY	181	QLRLAKWIDGFAGALFAGFIHLIITSR	206			
DB	181	QLRLAKWIDGFAGALFAGFIHLIITSR	206			
<p>-----</p>						
RESULT	2					
ID	YD07_HAEIN					
CD	Q57320; O05057;	STANDARD;	PRT;	210 AA.		
DT	01-NOV-1997 (Rel. 35, Created)					
DT	01-NOV-1997 (Rel. 35, Last sequence update)					
DT	15-JUL-1998 (Rel. 36, Last annotation update)					
DE	HYPOTHETICAL PROTEIN H11307.					
GN	H11307					
OS	Haemophilus influenzae.					
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;					
QC	Haemophilus					
OX	NCBI_Taxid=727;					
RN	[]					
RP	SEQUENCE FROM N.A.					
RC	STRAIN-RD / KW20 / ATCC 51907;					
RX	MEDLINE=95350650; PubMed=542800;					
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,					
RA	McKenney K., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,					
RA	Scott J.D., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,					
RA	Weidman J.F., Shirley R.A., Spriggs T., Hedblom E., Cotton M.D.,					
RA	Utterback T.T., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,					
RA	Fine L.D., Frichman J.L., Fuhrmann J.U., Geoghegan N.S.M.,					
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,					
RA	Venter J.C.;					

```

RA Duncan M., Allen E., Araujo R., Aparicio A.M., Chung E., Davis K.,
RA Federspiel N., Hyman R., Kalman S., Komp C., Kurdi O., Lew H.,
RA Lin D., Namath A., Oefner P., Roberts D., Schramm S., Davis R.W.;
RA Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: AE000140; AAC73431.1; -.
DR EMBL: U73857; AAB18053.1; -.
DR ECGene; EG13598; yahn.
DR InterPro; IPR001123; -.
DR Pfam; PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 22 42 POTENTIAL.
FT TRANSMEM 59 79 POTENTIAL.
FT TRANSMEM 85 105 POTENTIAL.
FT TRANSMEM 164 184 POTENTIAL.
FT CONFLICT 213 223 LRLIEGVTVR -> YA (IN REF. 2).
SQ SEQUENCE 223 AA; 24811 MW; E17F5ABC31EE3F26 CRC64;

Query Match 24.8%; Score 261.5; DB 1; Length 223;
Best Local Similarity 28.2%; Pred. No. 3e-14;
Matches 50; Conservative 50; Mismatches 83; Indels 15; Gaps 3;

QY 4 LFLVAMVHVALMSGPDFFVVSOTAVSRKEAMGVLTGTCGVMMWAGIALIGLHLI 63
:||||: ::|||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
DB 20 VLTAVGL-FVITFENPGANLFFVVTSLASGRAGVLTGLGVALGDAFYSGILGFLATL 78
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
QY 64 IEKMAWLTILMVGGLYLCHWGYQMLRGALKKAVSAPAPQVELAKGRS-----FLK 117
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
DB 79 ITCQEEIFSLIRIVGGAYLLNFAWCSMR-----RQSTPQMSTLQQIPAPWVFFRR 130
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
QY 118 GLLTNLNPKAIIFGVSFLVGDVNTTARWGIKIFALIVETLAWFTVVASLFAIPQMR 177
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
DB 131 GLITDLSNPQTVLFFSIFSVTLNAETPTWRLMAWAGIVLASIIRWFVLSQATSLPAVR 190
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
QY 178 RYQORLAKWIDGPAALPAGFGIHLI 203
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
DB 191 RAYGNQWVASRVIGAIIGVFALRI 216
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

RESULT 4
YBF7_PSEAE STANDARD; PRT; 216 AA.
AC P38102; Q9HV45;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-2000 (Rel. 40, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL PROTEIN PA4757.
GN PA4757.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OC NCBI_TaxID=287;
[1]
RN
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U04992; AAA19047.1; -.
DR EMBL: U81259; AAB39251.1; -.
DR EMBL: AE000489; AAG08143.1; -.
DR InterPro; IPR001123; -.
DR Pfam; PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 48 68 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 134 154 POTENTIAL.
FT TRANSMEM 156 176 POTENTIAL.
FT TRANSMEM 191 211 POTENTIAL.
FT CONFLICT 109 109 A -> G (IN REF. 1).
FT CONFLICT 119 119 P -> R (IN REF. 1).
FT CONFLICT 122 122 K -> Q (IN REF. 1).
SQ SEQUENCE 216 AA; 23248 MW; A30A08E714591B8D CRC64;

Query Match 19.6%; Score 207; DB 1; Length 216;
Best Local Similarity 29.4%; Pred. No. 6.2e-10;
Matches 64; Conservative 30; Mismatches 88; Indels 36; Gaps 4;

QY 5 ELTVAMVHVALMSGPDFFVVSOTAVSRKEAMGVLTGTCGVMMWAGIALIGLHLI 64
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
DB 9 FWTYVLGVVFIILPGNSLFLVLTSAQRGATGYRACGVFLGDAVLMLLSALGVASIL 68
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
QY 65 EKMAWLTILMVGGLYLCHWGYQMLRGALKK-----EAVSAPAPQVELAKGRSPLKGL 120
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
DB 69 KAEPMLFGLKALYALFYLVGVGMURGAWRLRNPNPATAQAQEDV---HQPFRKALL 125
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
QY 121 TNLANKAIIFGVSFLVGDVNTTARWGIKIFALIVETL-----AMF 164
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
DB 126 LLSLNPKAILEFFISFFIQFVDPGYAYPGLSFLVLAIVLELVSALYSFLITGVRLAAWF 185
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
QY 165 TVVASLFPQMRRGYQRLAKWIDGPAALPAGFGIHL 202
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:
DB 186 -----RRQRLAAGATSGVGLFVFGVKL 210
:||||: |||: |||: |||: |||: |||: |||: |||: |||: |||: |||:

RESULT 5
YEAS_ECOLI STANDARD; PRT; 212 AA.
ID YEAS_ECOLI
AC P76249; O07971; O07969;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 23.2 KDA PROTEIN IN GAPA-RND INTERGENIC REGION.
GN YEAS.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;

```

```
OC Escherichia.
OX NCBI_TaxID=562;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL science 277:1453-1474(1997).
RN
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97251358; PubMed=9097040;
RA Itoh T., Aiba H., Baba T., Fujita K., Hayashi K., Inada T., Isono K.,
RA Kasai H., Kimura S., Kitakawa M., Kitagawa M., Makino K., Miki T.,
RA Mizubuchi K., Mori H., Mori T., Motomura K., Nakade S., Nakamura Y.,
RA Nashimoto H., Nishio Y., Oshima T., Saito N., Sampei G., Seki Y.,
RA Sivasubram S., Tagami H., Takeda J., Takemoto K., Wada C.,
RA Yamamoto Y., Horiuchi T.;
RT "A 460-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 40.1-50.0 min region on the linkage map.";
RL DNA Res. 3:379-392(1996).
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC 1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; AE000274; AAC74868.1; -
DR EMBL; D90823; BAA15593.1; -
DR EMBL; D90824; BAA15602.1; -
DR EcoGene; EGI3505; Yeas.
DR InterPro; IPR001123; -.
DR Pfam; PF01810; Lyse; 1.
DR Hypothetical protein; Transmembrane.
KW TRANSMEM 12 32 POTENTIAL.
FT TRANSMEM 49 69 POTENTIAL.
FT TRANSMEM 71 91 POTENTIAL.
FT TRANSMEM 122 142 POTENTIAL.
FT TRANSMEM 153 173 POTENTIAL.
FT TRANSMEM 188 208 POTENTIAL.
SQ SEQUENCE 212 AA; 23200 MW; 342E0DF348C9AD9A CRC64;

Query Match 18.0%; Score 190; DB 1; Length 212;
Best Local Similarity 28.2%; Pred. No. 1,4e-08;
Matches 57; Conservative 36; Mismatches 85; Indels 24; Gaps 6;

QY 13 IVALMSPGDFFVVSQTAVERSRKEAMGVLTGCG-----VWVAGIALGLHLIEKM 67
DB 18 IFIVLPGDNTFLVKNSVSGMKGGYLACGVFGDAVLMFLAWAGVATL-----IKTT 72
QY 68 AMLHTLIMVGGGLYLCWVGQMLRGALKKEAVSAPAPQVEIAKSGRSLKGLTNLNPK 127
DB 73 PLFLNIVRYLGAFLYLLGSKILYATLKGNSEAKSDPEQV---GAIFKRLILSLTNPK 129
QY 128 AIYFGSVFSLFVGDVNGVTTARWGIFALIVETLAW-----FTVASLFLPQKRRGY 180
DB 130 ALLFYVSPFVQFDVNAFHT---GISFILAATLELVFCVLSFLIISGAF-VTQYIRTK 185
QY 181 ORLAKWIDGCFAGALFAGFGIHL 202
DB 186 KKLAKVGNLSLGLMFGVGAARL 207

RESULT 6
YRHP-BACSU STANDARD; PRT; 210 AA.
ID
AC 005406; 1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 23.4 KDA PROTEIN IN AAPA-SIGV INTERGENIC REGION.
GN YRHP.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1423;
RN
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RA Kunst F., Ogasawara N., Yoshikawa H., Danchin A.;
RL Submitted (NOV-1997) to the EMBL/GenBank/DBJ databases.
RN
RP SEQUENCE OF 1-141 FROM N.A.
RC STRAIN=168;
RX MEDLINE=97453479; PubMed=9308178;
RA Sorokin A., Bolotin A., Purnelle B., Hilbert H., Lauber J.,
RA Duesterhoeft A., Ehrlich S.D.;
RT "Sequence of the Bacillus subtilis genome region in the vicinity of
the lev operon reveals two new extracytoplasmic function RNA
polymerase sigma factors Sigv and Sigz.";
RL Microbiology 143:2939-2943(1997).
CC 1- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC 1- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed, usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
DR EMBL; Z99117; CAB14652.1; -
DR EMBL; U93874; AAB80873.1; -
DR Subtilist; BG12304; Yrhp.
DR InterPro; IPR001123; -.
DR Pfam; PF01810; Lyse; 1.
DR Hypothetical protein; Transmembrane.
KW TRANSMEM 5 25 POTENTIAL.
FT TRANSMEM 50 70 POTENTIAL.
FT TRANSMEM 75 95 POTENTIAL.
FT TRANSMEM 155 175 POTENTIAL.
SQ SEQUENCE 210 AA; 23389 MW; 2A37D9419FDB0A58 CRC64;

Query Match 17.1%; Score 180.5; DB 1; Length 210;
Best Local Similarity 24.5%; Pred. No. 7,7e-08;
Matches 51; Conservative 41; Mismatches 101; Indels 15; Gaps 5;

QY 2 LMLFTVAMVHIVAMSPGDDFFVVSQTAVERSRKEAMGVLTGCGVWVAGIALGLH 61
DB 4 LLAIPIDAMVWII---PGADTLMVKNVIRVGPAGRNILGLATGLSPWVAILGLS 60
QY 62 LIIEKMWLHTLIMVGGGLYLCWVGQMLRGALKKEAVSAPAPQVE---LAKSGR----- 113
DB 61 VVIKSVILFTIKYLGAAALYILG---VKSFSAKSMFLSDDDMSQAKNWSPPKRYKT 117
QY 114 SFLKGLTLNLANPKAIYFGSVFSLFVGDVNGVTTARWGIFALII-VETLAWFTVASLFA 172
DB 118 SPMQGSLSNLPKTVIVVYVTIMPFQINLNGINQQLILASILTLLAVLWFLFLVYIID 177
QY 173 LPQMRGQYRLAKWIDGCFAGALFAGFGI 200
DB 178 YAKKWKNSKFKQVFKITGILVGFGI 205
```

```

RESULT 7
CHPE_PSEAE
ID CHPE_PSEAE STANDARD; PRT; 203 AA.
AC O87005;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE CHEMOTACTIC TRANSDUCTION PROTEIN CHPE.
OS CHPE OR PA0417.
GN Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RA Whitchurch C.B., Young M.D., Hobbs M., Mattick J.S.;
RT "Pseudomonas aeruginosa chemotactic transduction genes pill, chpA
RT chpB and downstream genes chpC, chpD and chpE.";
RT Submitted (Nov-1996) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Gaber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Lardig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RN [3]
RL Nature 406:959-964 (2000).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; U79580; AAC23935.1; -
DR EMBL; AE004479; AAG03806.1; -
DR InterPro: IPR001123; -
DR Pfam: PF01810; Lyse; 1.
KW Transmembrane.
FT TRANSMEM 3 23 POTENTIAL.
FT TRANSMEM 45 66 POTENTIAL.
FT TRANSMEM 69 89 POTENTIAL.
FT TRANSMEM 123 143 POTENTIAL.
FT TRANSMEM 149 169 POTENTIAL.
FT SEQUENCE 203 AA; 21290 MW; 195553C048AAD099 CRC64;
Query Match 15.1%; Score 159; DB 1; Length 203;
Best Local Similarity 27.7%; Pred. No. 3.8e-06;
Matches 59; Conservative 38; Mismatches 86; Indels 30; Gaps 8;
QY 1 MMLFLTVAMVHIVALMSFGDPFFVTSOTAVSRSRKEAMMGVLGITCGVWVWAGIALGL 60
DB 1 MLIIFLAALFGFAFNVSPGVFSETLRGLTGGRFPALLVQLGSLIGDANVALLGLTGL 60
QY 61 HLII--EKMAWLHTLMVGGGLYLCWGYQMLRGALKKEAVSAPAPQVELAKSGR-SPLK 117
DB 61 ALLLGEQVRIPUTLACAA--YLAWLGVQGLR-----DAWSPPLAEDAGEGQGNAPGA 112
QY 118 GLLTNLNANPAIIYFGSVFLVGVGNVGT---TARWGIFALIIVETLAWFTVWASLFLALP 174
DB 113 GAATSLSNPKNVVYWCALGSLAGSALAGIVDCTPNQAQSLVFFAGFMLSLWCFCAAL--VD 170

```

```

QY 175 QMRG-----YQRLAKWIDGFPAGALFAGFGIHLI 203
DB 171 WLARNTSLFWHRVS-----YAGCGVLLL 193
RESULT 8
YQ27_SYNY3
ID YQ27_SYNY3 STANDARD; PRT; 206 AA.
AC P74343;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 22.0 KDA PROTEIN SLR1627.
GN SLR1627.
OS Synechocystis sp. (strain PCC 6803).
OC Bacteria; Cyanobacteria; Chroococcales; Synechocystis.
OX NCBI_TaxID=1148;
RN [1]
RP SEQUENCE FROM N.A. PubMed=8905231;
RX MEDLINE=97061201;
RA Kaneko T., Sato S., Kotani H., Tanaka A., Asamizu E., Nakamura Y.,
RA Miyajima N., Hiroseawa M., Sugiyama M., Sasaki M., Kimura I.,
RA Rosouchi I., Matsuno A., Muraki A., Nakazaki N., Naruo K.,
RA Okumura S., Shimpo S., Takeuchi C., Wada T., Watanabe A.,
RA Tamada M., Yasuda M., Tabata S.;
RT "Sequence analysis of the genome of the unicellular cyanobacterium
RT Synechocystis sp. strain PCC6803. II. Sequence determination of the
RT entire genome and assignment of potential protein-coding regions.";
RN [2]
RL DNA Res. 3:109-136 (1996).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
-----
EMBL; D90914; BAA18437.1; -
DR InterPro: IPR001123; -
DR Pfam: PF01810; Lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 9 29 POTENTIAL.
FT TRANSMEM 47 67 POTENTIAL.
FT TRANSMEM 74 94 POTENTIAL.
FT TRANSMEM 150 170 POTENTIAL.
FT SEQUENCE 206 AA; 22041 MW; EF493754B8F264AF CRC64;
Query Match 15.1%; Score 159; DB 1; Length 206;
Best Local Similarity 26.8%; Pred. No. 3.9e-06;
Matches 56; Conservative 39; Mismatches 96; Indels 18; Gaps 6;
QY 3 MFLFTVMVHIVALMSFGDPFFVTSOTAVSRSRKEAMMGVLGITCGVWVWAGIALGLHL 62
DB 9 ILSLFAMLIILAAAL--PSLSVLTAVSSKASGGFIHGLFAALGVVLGDIIFILIALWGLAF 66
QY 63 IEKMAWLHTLMVGGGLYLCWGYQMLRGALKKEAVSAPAPQVELAKSGRSLKGLTN 122
DB 67 LEGAMGDFEVILKVIISGYLSWLGIINTIRAKVNNOSLA----KVDVKSLSFSFAGLIT 122
QY 123 LANPKAIYFGSVFLVGVGNVGTARWGIFALIIVETLAWFTVWASLFLALPQMRG-- 180
DB 123 LADKAVFLYGLFTFTV--DVNNIAYLDI-AVILITAILTVGGVKIFYAFLAHRGLLI 179
QY 181 ---ORLAKWIDGFGALFAGFGIHLIIS 205
DB 180 SRQNKRNLYL---AGALMISVGVLLIS 205
RESULT 9

```

```
RHTB_ECOLI
ID RHTB_ECOLI STANDARD; PRT; 206 AA.
AC P27847;
DT 01-AUG-1992 (Rel. 23, Created)
DT 15-AUG-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HOMOSERINE/HOMOSERINE LACTONE EFFLUX PROTEIN.
GN RHTB
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=92358234; PubMed=1379743;
RA Daniels D.L., Plunkett G. III, Burland V.D., Blattner F.R.;
RT "Analysis of the Escherichia coli genome: DNA sequence of the region
RL from 84.5 to 86.5 minutes.";
RL Science 257:771-778(1992).
RN [2]
RP CONCEPTUAL TRANSLATION.
RA Rudd K.E.;
RL Unpublished observations (JUN-1999).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=99313167; PubMed=10386596;
RA Zakataeva N.P., Aleshin V.V., Tokmakova I.L., Troshin P.V.,
RA Livshits V.A.;
RT "The novel transmembrane Escherichia coli proteins involved in the
RT amino acid efflux";
RL FEBS Lett 452:228-232(1999).
CC -!- FUNCTION: CONDUCTS THE EFFLUX OF HOMOSERINE AND HOMOSERINE
CC LACTONE.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; A FRAMESHIFT HAD TO
CC BE INTRODUCED IN POSITIONS 60 TO EXTEND THIS ORF.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M87049; AAA67620.1; ALT_FRAME.
CC PIR; S30714; S30714.
CC EcoGene; EG11469; rhtB.
CC InterPro; IPR001123; -.
CC Pfam; PF01810; LysE; 1.
CC Transport; Transmembrane.
CC FT TRANSMEM 5 25 POTENTIAL.
CC FT TRANSMEM 45 65 POTENTIAL.
CC FT TRANSMEM 68 88 POTENTIAL.
CC FT TRANSMEM 117 137 POTENTIAL.
CC FT TRANSMEM 148 168 POTENTIAL.
CC FT TRANSMEM 182 202 POTENTIAL.
CC SQ SEQUENCE 206 AA; 22427 MW; 1164F17738509C8C CRC64;

Query Match 13.7%; Score 144.5; DB 1; Length 206;
Best Local Similarity 20.7%; Pred. NO. 5.5e-05;
Matches 43; Conservative 44; Mismatches 96; Indels 25; Gaps 4;

QY 10 MVHIVALSPGDDFFVSQTAVSRKSEAMMGVLGTCGVMWAGIALGLHLIERKAW 69
Db 11 L7SITLSLSPSGGAINWTTSLNHGYRGAVASIAQTGLAIHIVGVGLGTLFSRSVI 70
QY 70 LHTLMVGGGLYCWGMQYMLR--GALKKEAVSAPAPQVELAKSGRSFLKGLLTNLNPK 127

Db 71 AFEVLKWAAGAAVLIWGIQWRAAGAI DLKSLASTOSRRHL-----FQRAVFNLTNPK 124
QY 128 AIIYFGSVFSLFVGDVNGVTARMGIFALIIVETLAWFTVVASLFAIPOMRRGYQRLAKWI 187
Db 125 SIVFLAALFPQFIMPQPOLMOYIVLGV-----TIVVDIIVMIGYATLAQRIALWI 176
QY 188 DG-----PAGALFAGFGIHLIISR 206
Db 177 KGPQKMKALNKIFGSLFPLVGLLASAR 204

RESULT 10
YGA_AERSA STANDARD; PRT; 206 AA.
AC P70775;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 21.5 KDA PROTEIN IN ASAR-CDPD INTERGENIC REGION.
GN YGA.
OS Aeromonas salmonicida.
OC Bacteria; Proteobacteria; gamma subdivision; Aeromonadaceae;
OX NCBI_TaxID=645;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-NCIMB 1102;
RA Swift S., Karlyshev A.V., Fish L., Durant E.L., Winson M.K.,
RA Williams P., MacIntyre S., Stewart G.S.A.B.;
RT Submitted (SEP-1996) to the EMBL/GenBank/DBJ databases.
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE LYSF/YGA FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; U65741; AAB70019.1; ALT_INIT.
CC InterPro; IPR001123; -.
CC Pfam; PF01810; LysE; 1.
CC KW Hypothetical protein; Transmembrane.
CC FT TRANSMEM 1 21 POTENTIAL.
CC FT TRANSMEM 37 57 POTENTIAL.
CC FT TRANSMEM 65 85 POTENTIAL.
CC FT TRANSMEM 116 136 POTENTIAL.
CC FT TRANSMEM 148 168 POTENTIAL.
CC FT TRANSMEM 185 205 POTENTIAL.
CC SQ SEQUENCE 206 AA; 21505 MW; D1C2C492CDA0179A CRC64;

Query Match 10.6%; Score 111.5; DB 1; Length 206;
Best Local Similarity 26.0%; Pred. NO. 0.023;
Matches 52; Conservative 32; Mismatches 91; Indels 25; Gaps 11;

QY 14 VALMSP-GDFFVSQTAVSRKSEAMMGVLGTC-----GYVWAGIALGLHLI-IEK 66
Db 13 LAMITPIGAQNAPVLSRGIHRNH-HLLAATLCCLCLILIGIVFGGANLAAASPIGAL 71
QY 67 MAWLHPLIMVGGGLYCWGMQYMLRGAKE-AYSAPAPQVELAKSGRSFLKGLLTNLN 125
Db 72 LTV-----GGVFLCFWGRSRSWAGOGAALADSPRLMGVRSVLAITLGV--TLIN 122
QY 126 PKAIIYFGSVFSL-FVGDVNGVTARMGIFALIIVETLAWFTVVA--SLFALPOMRRGYQR 182
Db 123 PH--VYLDPLMLGLSFGSQFAEPLRPAFAAGAMLASLVWFYSFAFGAALSPWLARG--R 178
QY 183 LAKWIDGFAGALFAGFGIHL 202
Db 179 VQAIDTVIGLIMLGLALQL 198
```


[illegible]

01-OCT-1996 (Rel. 34, Last sequence update)
15-JUL-1998 (Rel. 36, Last annotation update)
DE HYPOTHETICAL 23.2 KDA PROTEIN IN SRM-FBA INTERGENIC REGION (ORF 5).
GN YGGA.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdlvision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE OF 15-211 FROM N.A.
RC STRAIN=K12 / CS520;
RX MEDLINE=89313302; PubMed=2546007;
RA Alefounder P.R., Baldwin S.A., Perham S.A., Short N.J.;
RT "Identification, molecular cloning and sequence analysis of a gene
RT cluster encoding the class II fructose 1,6-bisphosphate aldolase, 3-
RT phosphoglycerate kinase and a putative second glyceraldehyde 3-
RT phosphate dehydrogenase of Escherichia coli.";
RL Mol. Microbiol. 3:723-732(1989).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (PROBABLE).
CC -!- SIMILARITY: BELONGS TO THE LYSE/YGGA FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; U28377; AAA69090.1; -;
DR EMBL; AE000375; AAC75960.1; -;
DR EMBL; X14436; CAA32607.1; -;
DR PIR; S04736; Q0EC5A.
DR EcoGene; EG11159; ygga.
DR InterPro; IPR001123; -;
DR Pfam; PF01810; lyse; 1.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 1 21 POTENTIAL.
FT TRANSMEM 37 57 POTENTIAL.
FT TRANSMEM 68 88 POTENTIAL.
FT TRANSMEM 111 131 POTENTIAL.
FT TRANSMEM 147 167 POTENTIAL.
FT TRANSMEM 179 199 POTENTIAL.
FT SEQUENCE 211 AA; 23175 MW; 2DAFE27B6A9BE822 CRC64;

Query Match 9.8%; Score 103; DB 1; Length 211;
Best Local Similarity 27.8%; Pred. No. 0.11;
Matches 54; Conservative 25; Mismatches 22; Gaps 9;
QY 4 LELTAVMHHVIALMSPGDFFVFSQTAVSRSKEAMGVLTGCVWVWAGIALLGLHLI 63
DB 9 LALGAAMI-----LPLGQNAFVNMQRIQQYHIMIALICASDLVLCAGI--FGGSAL 61
QY 64 IEFMAWHLTLWGGGLYLCWGMQMLRGALKKEAVSAPAPQVELAKSGRFLKGLTLNL 123
DB 62 LMQSPWLLALVTWGVAFLLWYGFAGFKTAMSSNIELASA---EYMKQGR--WKIATML 116
QY 124 A----NPKAIYFGSVSLFVGDNVGTATRWGIFAL--IIVETLAWFTVVASLFA--LPQM 176
DB 117 AVTLNPNHYLDTFWLGSGLQGLDVEPKRW--FALGTISASFLWFFGLAALLAALAPRL 174
QY 177 RRG-YORLAKWIDG 189

DB 175 RTAKAQRILNVLVG 188

RESULT 14
YFIK_ECOLI
ID YFIK_ECOLI STANDARD; PRT; 195 AA.
AC P38101;
DT 01-OCT-1994 (Rel. 30, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE HYPOTHETICAL 21.2 KDA PROTEIN IN SRMB-UNG INTERGENIC REGION.
GN YFIK.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdlvision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
RN [5]
RP IDENTIFICATION.
RA Rudd K.E.;
RA Unpublished observations (AUG-1994).
CC -!- SUBCELLULAR LOCATION: INTEGRAL MEMBRANE PROTEIN (POTENTIAL).
CC -!- SIMILARITY: BELONGS TO THE RHT FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL; D13169; -; NOT ANNOTATED CDS.
DR EMBL; D64044; -; NOT ANNOTATED CDS.
DR EMBL; AE000344; AAC75631.1; -;
DR EMBL; D90886; BAA16464.1; -;
DR EcoGene; EG12445; yfik.
KW Hypothetical protein; Transmembrane.
FT TRANSMEM 8 28 POTENTIAL.

Db 121 TNLNPKAIYFGSVFSLVFGDVGAAARWGIFALITLLETAWFTVVASLFPALPKMRGGY 180
QY 181 ORLAKWIDGFAGALPAGRGHILIIISR 206
Db 181 QRLAKWIDGFAGALPAGRGHILIIISR 206
RESULT 2
QYKVF5 PRELIMINARY; PRT; 222 AA.
AC Q9KVF5;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE HYPOTHETICAL PROTEIN VC0191.
GN VC0191.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae";
RL Nature 406:477-483(2000).
DR EMBL; AB004109; AAF93367.1; -.
DR TIGR; VC0191; -.
SQ SEQUENCE 222 AA; 23485 MW; B5B83AC3804E5E71 CRC64;

Query Match 29.7%; Score 313.5; DB 2; Length 222;
Best Local Similarity 31.3%; Pred. No. 1.3e-18;
Matches 62; Conservative 52; Mismatches 79; Indels 5; Gaps 2;
QY 2 LMLFLTVAMVHIVALMSGPDPFFVQSQTAVSRSRKEAMGVLGITCGVMVWAGIALGLHL 61
Db 4 MSLATAGVHETALLSPGPDVALYVQNAQTGGRKTGVNIALGLSCGILVHLILSLSGIS 63
QY 62 LIIEKMAWLHTLIMVGGGLYLCWMGYQMLRGALKKEAVS---APAPQVELAKSGRFLK 117
Db 64 YLVKQPMFLNLLQLAGGSYLLYLGCAGALQSVMAQKNASTPTHTSPAPSI-LGNRRQAFK 122
QY 118 GLLTNLNLANPKAIYFGSVFSLVFGDVGNTTARWGIFALIIIVETLAWFTVVASLFPALPQMR 177
Db 123 GMTNLNLANPKAIYFGSVFSLVFGDVGNTTARWGIFALIIIVETLAWFTVVASLFPALPQMR 182
QY 178 RGYORLAKWIDGFAGALF 195
Db 183 QRMQRTSRVSDSCAAVF 200

RESULT 3
QYK8S1 PRELIMINARY; PRT; 210 AA.
ID Q9K8S1;
AC Q9K8S1;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-OCT-2000 (TrEMBLrel. 15, Last annotation update)
DE BH2332 PROTEIN.
GN BH2332.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]

RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA Takami H., Nakasone K., Takaki Y.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL; AP001517; BAB06651.1; -. C49C7DE7C86F6F75 CRC64;
SQ SEQUENCE 210 AA; 23712 MW;

Query Match 27.7%; Score 292; DB 2; Length 210;
Best Local Similarity 32.4%; Pred. No. 7.3e-17;
Matches 67; Conservative 37; Mismatches 97; Indels 6; Gaps 2;
QY 4 LFLTVAMVHIVALMSGPDPFFVQSQTAVSRSRKEAMGVLGITCGVMVWAGIALGLHLI 63
Db 1 MFVEVFIYGLLAGMSGPDPFFIYVKNKSLGFGARVGLTSLGTASALIVHIYTVVLGFAFL 60
QY 64 IERMAWLHTLIMVGGGLYLCWMGYQMLRGA--LKKDAVSAPAPQVELAKSGRS----FLK 117
Db 61 IETYPALFFITQLLGAAYLWLGFHAIIRSPPKKEAEIETQPIQSTKDKSSIQSGKE 120
QY 118 GLLTNLNLANPKAIYFGSVFSLVFGDVGNTTARWGIFALIIIVETLAWFTVVASLFPALPQMR 177
Db 121 GFITNLNLANPKAALEFLSIFSOFTIPQTADVWRMYGLEVVAVGLWESFLAIFISYKHFR 180
QY 178 RGYORLAKWIDGFAGALPAGFGHILII 204
Db 181 RFYQTHSYWFDRLGAAALLFFAIRIII 207

RESULT 4
QYRIF0 PRELIMINARY; PRT; 212 AA.
ID Q9RIF0
AC Q9RIF0; 2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2000 (TrEMBLrel. 14, Last annotation update)
DE YFUD PROTEIN.
GN YFUD.
OS Yersinia enterocolitica (type 0:8).
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Yersinia.
OX NCBI_TaxID=34054;
RP SEQUENCE FROM N.A.
RC STRAIN=WA-314;
RA Saken E.M.;
RL Submitted (JAN-1995) to the EMBL/GenBank/DBJ databases.
DR EMBL; Z47200; CAB61500.1; -.
DR INTERPRO; IPR001123; -.
DR PFAM; PF01810; LysE; 1.
SQ SEQUENCE 212 AA; 22790 MW; 0C18DA690C479702 CRC64;

Query Match 25.6%; Score 269.5; DB 2; Length 212;
Best Local Similarity 28.1%; Pred. No. 5.3e-15;
Matches 59; Conservative 52; Mismatches 80; Indels 19; Gaps 4;
QY 3 MLFLTVAMVHIVALMSGPDPFFVQSQTAVSRSRKEAMGVLGITCGVMVWAGIALGLHL 62
Db 6 VIITLIGL-FVLTFFINPGANLLVVQTSLSGKNAGLLTGLGALGDALYSGLGFMAA 64
QY 63 IIEKMAWLHTLIMVGGGLYLCWMGYQMLR-----GALKKEAVSAPAPQVELAKSGR 113
Db 65 LTAEAGGALFSAIKIGGLYLVWYAYNKNVRHRQELHMGMAV---ATSSITPWYVF----- 116
QY 114 SFLKGLLTNLNLANPKAIYFGSVFSLVFGDVGNTTARWGIFALIIIVETLAWFTVVASLFPAL 173
Db 117 -FRRLTDLNPNQTVLFFISFVTLTPTTPANAKWALGIIVASILWRSLLSMAFSR 175
QY 174 PQMRRGYORLAKWIDGFAGALFAGFGHILI 203
Db 176 PAVRRAYGVQVHLLSGIIGVAVGAFGLRLI 205

```
RESULT 5
Q9L6H7 PRELIMINARY; PRT; 212 AA.
AC Q9L6H7;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE AMINO ACID EFFLUX-LIKE PROTEIN.
OS Brucella melitensis.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Brucellaceae; Brucella.
OX NCBI_TaxID=29459;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=16M;
RA Zygmunt M.S., Diaz M.A., Teixeira-Gomes A.P., Cloeckaert A.;
RT "Cloning, nucleotide sequence, and expression of the Brucella
RT melitensis sucB gene coding for a dihydrolipoamide succinyltransferase
RT homologous protein.";
RL Submitted (FEB-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF235020; AAF43702.1; -; OC21287CER665F56 CRC64;
SQ SEQUENCE 212 AA; 23158 MW; 0C21287CER665F56 CRC64;

Query Match 23.9%; Score 251.5; DB 2; Length 212;
Best Local Similarity 29.9%; Pred. No. 1.6e-13;
Matches 59; Conservative 41; Mismatches 88; Indels 9; Gaps 3;

Qy 11 VHIWALMSPGDPFFVFSOTAVSRKRAMMVGITCGVMWAGIALGLHLIEKMAWL 70
Db 14 IFVFAIVSPGADLANVIRQSLHGREAITSEGITATLMMHVTVVLGLIISRYL 73
Qy 71 HTLWVGGLYLCWGMQMLR-GALKKEAVSAPAPQVEALAKSGRSFLK----GLTNTLAN 125
Db 74 FNIWKGCWGANLVYIGFKALRACTTKEA---GPDLKEPRKQGFKAFLGFGFAANALN 129
Qy 126 PKAIYFGSVFSLFVGNVGTARWGIFALIIVETLAWTFVVASLFAIPOMRRGYORLAK 185
Db 130 PKAVFFSLFSTVHVHTTEVKLGVGMATALLISWFGVSVFFMTTPKRAAFSRASK 189
Qy 186 WIDGFAGALFAGFIHL 202
Db 190 WIDRTSGVWFIALGLKL 206

RESULT 6
Q9KFP9 PRELIMINARY; PRT; 207 AA.
AC Q9KFP9;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DE BH0429 PROTEIN.
GN BH0429.
OS Bacillus halodurans.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=C-125 / JCM 9153;
RA Takami H., Nakasone K., Takaki Y.;
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP001508; BAB04148.1; -;
SQ SEQUENCE 207 AA; 22317 MW; 9FED148C8E8DD3F0 CRC64;

Query Match 19.1%; Score 201; DB 2; Length 207;
Best Local Similarity 28.3%; Pred. No. 2.3e-09;
Matches 63; Conservative 39; Mismatches 81; Indels 40; Gaps 9;

Qy 2 LMLFLTVMVHIVALMSPGDPFFVFSOTAVSRKRAMMVGITCGVMWAGIALGLH 61
```

```
Db 6 LLSFLGVA---VLLTLPMPDILFLVLAQSMSONRQAGIVTALGLCTGLLVHTAAVVGIS 62
Qy 62 LIIEKMAWLHTLIWVGSG--LYLCWGMGYOMLRGALKKEAVSAPAPQVEALAKSGRSFLK 118
Db 63 ALIYOSALAFVVKYAGAAAYLLYLAWRAFQ-----KGEGLSID-KOTTUA-YGALYKKG 115
Qy 119 LITNLNPKAIYFGSVFSLFVGNVGTARWG-----IPALIIVETLAWTF----165
Db 116 IIMVNLNPKVSLFFLALLPQFVNSGAG-SAPQWMLLGLLVGVFLIQAFITFSLVSWFAEKVG 174
Qy 166 --VVASLFAIPOMRRGYORLAKWIDGFPAGALFAGFIHLIISR 206
Db 175 QLLMRSSFIKNQHR-----YKGLLALIGLQVAFSK 206

RESULT 7
Q9KLA0 PRELIMINARY; PRT; 204 AA.
AC Q9KLA0;
DT 01-OCT-2000 (TREMBlrel. 15, Created)
DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
DT 01-OCT-2000 (TREMBlrel. 15, Last annotation update)
DE HYPOTHETICAL PROTEIN VCA0846.
GN VCA0846.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOP N16961 / SEROTYPE O1;
RX MEDLINE=2046933; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser G.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL: AE004412; AAF96744.1; -;
DR TIGR: VCA0846; -;
SQ SEQUENCE 204 AA; 21719 MW; D09887299659FBBD CRC64;
```

```
Query Match 18.7%; Score 197.5; DB 2; Length 204;
Best Local Similarity 25.2%; Pred. No. 4.4e-09;
Matches 51; Conservative 46; Mismatches 102; Indels 3; Gaps 1;

Qy 2 LMLFLTVMVHIVALMSPGDPFFVFSOTAVSRKRAMMVGITCGVMWAGIALGLH 61
Db 3 LTVWLSLFTICILGAMSPGSLAMVAKHSLAGRKNGFAAAWAHAFGIVYAFITLIGLA 62
Qy 62 LIIEKMAWLHTLIWVGGLYLCWGMGYOMLRGALKKEAVSAPAPQVEALAKSGRSFLKGLIT 121
Db 63 VVLRHQSPLVFKTISYAGAAAYLAYLGNALR---SKGGVAAKLESGESVSWQASREGLLI 119
Qy 122 NLANPKAIYFGSVFSLFVGNVGTARWGIFALIIVETLAWTFVVASLFAIPOMRRGYQ 181
Db 120 SLLSPKIALFFIALFSQYVAVGSDLTSKAAIIVITPLVVDGLWYSFITLILSSPRLLDKLR 179
Qy 182 RLAKWIDGFPAGALFAGFIHLI 203
Db 180 ARAVLIDRLSLGLVLAIRVL 201

RESULT 8
Q9RMX0 PRELIMINARY; PRT; 205 AA.
AC Q9RMX0;
DT 01-MAY-2000 (TREMBlrel. 13, Created)
```

```
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DE 01-JUN-2000 (Tremblrel. 14, Last annotation update)
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OG Bacillus anthracis.
OG Plasmid pX02.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OX NCBI_TaxID=1392;
RN [1]
RP SEQUENCE FROM N.A.
RA Okinaka R.T., Cloud K., Hampton O., Hill K.K., Keim P., Lamke G.,
RA Kuman R.T., Manter D., Martinez Y., Svensson R., Tatum L.R.,
RA Brown A.E., Jackson P.J.,
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF188935; AAF13668.1; -
DR INTERPRO: IPR001123; -
DR PFAM: PF01810; Lyse; 1.
KW Plasmid.
SQ SEQUENCE 205 AA; 22900 MW; 6DC91EE968B524B8 CRC64;

Query Match 18.0%; Score 190; DB 2; Length 205;
Best Local Similarity 26.1%; Pred. No. 1.9e-08;
Matches 53; Conservative 50; Mismatches 92; Indels 8; Gaps 6;

QY 2 LMLFTVAMVHVALMSPGDPFFVVSQTAVSRSRKEAMMGVLGITCGVMVWAGIALLGLH 61
Db 01-OCT-2000 (Tremblrel. 15, Last sequence update)
QY 3 LNIWITVLVGGIGVISPENWAVIKNSL-YSRSLGVSTVAGIATGSLIHIVYCLIG 61
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 62 LIIEKMAWLHTLMVGGGLYLCWGYQMLRGALKKEAVSAPQVELAKSGRSFLKLLT 121
Db 01-OCT-2000 (Tremblrel. 15, Last sequence update)
QY 62 VIISKSILLNTLKWIGVAYLLYIGIKLLRS--KKQSPAAIKKNE-STTWKAFRSGLT 118
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 122 NLANPKAIIFGVSFLFVGNDVGTARWGIFALI--VETLAWFTVVASLFPQMRRG 179
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 119 DMLNPKATFLYLAIFTQVIEPTNIFVQ-SVYGLTVMSVEIL-WHMLVFFLTHKSVRY 176
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 180 YORLAKWIDGFAGALFAGFGH 202
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 177 FLISHWIERVGTALILGIRL 199
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)

RESULT 9
Q9K775 PRELIMINARY; PRT; 208 AA.
AC Q9K775;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)
DT 01-OCT-2000 (Tremblrel. 15, Last annotation update)
DE DIHYDRODIPICOLINATE REDUCTASE.
GN BH3495.
OS Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus.
OX NCBI_TaxID=86665;
RN [1]
RP SEQUENCE FROM N.A.
RA Takami H., Nakasone K., Takaki Y.,
RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AP001519; BAB07214.1; -
SQ SEQUENCE 208 AA; 23158 MW; CAD7A46D63B8E45A CRC64;

Query Match 17.6%; Score 186; DB 2; Length 208;
Best Local Similarity 25.3%; Pred. No. 4e-08;
Matches 55; Conservative 49; Mismatches 83; Indels 30; Gaps 7;

QY 3 LMLFTVAMVHVALMSPGDPFFVVSQTAVSRSRKEAMMGVLGITCGVMVWAGIALLGLH 62
Db 01-OCT-2000 (Tremblrel. 15, Last sequence update)
QY 8 LLFMGSA---IMLIIPGDLVFTTQGTNGRKAGVITAMGLSIGNIVHTFLAAVIGLSL 64
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 63 IIEKMAWLHTLMVGGGLYLCWGYQMLRGALKKEAVSAPQVELAKSGRSFLKLLTN 122
```

```
Db 01-MAY-2000 (Tremblrel. 13, Last sequence update)
QY 65 IIQTSVVVFTIFKAGAWILFYLAALKH--RKEEIVNSEKTNLKG--LFLAGLINN 120
Db 01-JUN-2000 (Tremblrel. 14, Last annotation update)
QY 123 LANPKAIIFGVSFLFVG-DNVGTTARWGIFALIIVETLANFTVVASLFPQMRRGY- 180
Db 01-JUN-2000 (Tremblrel. 14, Last annotation update)
QY 121 VLNPKVAIFYLTFPQFVNVQSANISIQILGLIFIMTA---IIFSIF-----GYF 170
Db 01-JUN-2000 (Tremblrel. 14, Last annotation update)
QY 181 -----QRLAKWIDGFAGALFAGFGIHLIIS 206
Db 01-JUN-2000 (Tremblrel. 14, Last annotation update)
QY 171 SGTFRDLLKNSRFNEYMANIAATIIIGLGKLMYQ 207
Db 01-JUN-2000 (Tremblrel. 14, Last annotation update)

RESULT 10
Q9RSX2 PRELIMINARY; PRT; 241 AA.
AC Q9RSX2;
DT 01-MAY-2000 (Tremblrel. 13, Created)
DT 01-MAY-2000 (Tremblrel. 13, Last sequence update)
DT 01-JUN-2000 (Tremblrel. 14, Last annotation update)
DE CONSERVED HYPOTHETICAL PROTEIN.
GN DR1999.
OS Deinococcus radiodurans.
OC Bacteria; Thermus/Deinococcus group; Deinococcales; Deinococcus.
OX NCBI_TaxID=1299;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RI;
RX MEDLINE=20036896; PubMed=10567266;
RA White O., Eisen J.A., Heidelberg J.F., Hickey E.K., Peterson J.D.,
RA Dodson R.J., Haft D.H., Gwinn M.L., Nelson W.C., Richardson D.L.,
RA Mofat K.S., Qin H., Jiang L., Pamphile W., Crosby M., Shen M.,
RA Vamathevan J.J., Lam P., McDonald L., Utterback T., Zalewski C.,
RA Makarova K.S., Aravind L., Daly M.J., Minton K.W., Fleischmann R.D.,
RA Ketchum K.A., Nelson K.E., Salzberg S., Smith H.O., Venter J.C.,
RA Fraser C.M.;
RT "Genome Sequence of the Radioresistant Bacterium Deinococcus
radiodurans R1."
RL Science 286:1571-1577(1999).
DR EMBL: AE002037; AAF11548.1; -
DR TIGR: DR1999; -
DR INTERPRO: IPR000719; -
DR INTERPRO: IPR001123; -
DR PFAM: PF01810; Lyse; 1.
DR PROSITE: PS00107; PROTEIN_KINASE_ATP; UNKNOWN_1.
SQ SEQUENCE 241 AA; 25161 MW; A4017ABFEDB60374 CRC64;

Query Match 16.6%; Score 174.5; DB 2; Length 241;
Best Local Similarity 29.1%; Pred. No. 4.2e-07;
Matches 65; Conservative 37; Mismatches 80; Indels 41; Gaps 11;

QY 5 FLTVMVHVALMSPGDPFFVVSQTAVSRSRKEAMMGVLGITCGVMVWAGIALLGLHII 64
Db 01-OCT-2000 (Tremblrel. 15, Last sequence update)
QY 39 FLVAAV--LALL-PGCLMYILARSILGGRWAGIOSALGTGAGMVHVLASAVGLSALI 95
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 65 EKMAWLHTLMVGGGLYLCWGYQMLRGALKKEAVS-----APAPOVELAKSGRSFL 116
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 96 MASSLAFSVVKYAGAAYLYLG---LRVLSKEALSKESASLAAAP-----RQSELT 148
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 117 KGLLNLANPKAIIFGVSFLFVGNDVGTARWGIFALI-----IVETLAW 163
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 149 OGAMTELENPKATFLAVIPQFVKPATGHV--FGQFLLLGTTTSVVVNTINAMLVATLAG 206
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 164 FTVVASLFPQMRRGYORLAKWIDGFAGALFAGFGIHLIIS 206
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)
QY 207 P-LGARLOGNPFQNG-QKVAS-----GGAMIA-LGTYYAVER 241
Db 01-OCT-2000 (Tremblrel. 15, Last annotation update)

RESULT 11
Q9KS40 PRELIMINARY; PRT; 212 AA.
ID Q9KS40
AC Q9KS40;
DT 01-OCT-2000 (Tremblrel. 15, Created)
```



```
DR PFAM; PF01810; Lyse; 1.
SQ SEQUENCE 213 AA; 23614 MW; E97242DD507C8065 CRC64;

Query Match 15.5%; Score 163; DB 2; Length 213;
Best Local Similarity 25.3%; Pred. No. 3.3e-06;
Matches 47; Conservative 35; Mismatches 76; Indels 28; Gaps 5;

QY 2 LMLFLTVAMVHVALMSGPDPFFVVSQTAVSRKREAMGVLTGCGVMWVWAGIALGLH 61
: ||| | | |||| | | | | | | | | | | | | | | | | | | | | | | |
Db 19 IYFILTVLFMSI-----SPGPMAMFVLQSQKQKGVKTGLAAVLGTEIGVFYIVILTALGIS 74
QY 62 LIIEKMAWLHTLIMVGGGLYLCWGMQMLRGALKKEAVSAPQVELAKSGRSFLKGLLT 121
: : : : : : : | | | | | | | | | | | | | | | | | | | | | |
Db 75 TVLKEPSYITGLQIGRAYLLIYAL-----SWPQNASNQTPTASRSYTGTFIQGLI 130
QY 122 NLANKKAIIFYGVSFLFV--GDNVGTARNGIF-----ALII-----IVETL 161
|| ||| : : : | | | | | | | | | | | | | | | | | | | | | |
Db 131 NLTNPKRIVLFLSLIPQFVPRDSNAMTFVYGLIFNTSGLLVNFSVALLADRVNRMLSRV 190
QY 162 AWFTVV 167
Db 191 TWENYV 196

RESULT 14
Q9KJ3 PRELIMINARY; PRT; 208 AA.
AC Q9KJ3;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DE HYPOTHETICAL PROTEIN VCA0355.
GN VCA0355.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
QY 17 MSPGDPFFVVSQTAVSRKREAMGVLTGCGVMWVWAGIALGLHIIIEKMAWLHTLIWV 76
: ||| | | : : : | | | | | | | | | | | | | | | | | |
Db 18 LSPGSGAINTMTTSINHGYRGAVASIAQLQTGLGIHIVLVGVGLTFLSRSLAFELKW 77
QY 77 GGGLYLCWGMQMLR--GALKKEAVSAPQVELAKSGRSFLKGLLTNLANKKAIIFYGS 134
: ||| | | | | | | | | | | | | | | | | | | | | | |
Db 78 AGAAYLILWLGIOQWRAGAILDLHTLA-----QTQSRGRLEKRAIFVNLNPKSIYFLAA 131
QY 135 VFSLEFVGDNVGTARTARIGIFALIIIVETLAWFTVVASLFPQMRRGYQRLAKWIDG----- 189
: | | | | | | | | | | | | | | | | | | | | | | | |
Db 132 LFEPQIMPQPOLAQY----LILGVT----TIVYDMVVMVTGYATLAQRIAAWIKGPKQMR 183
QY 190 ----FAGALFAGFGIHLIISR 206
: | | | | | | | | | | | | | | | | | | | | | |
Db 184 ALNKAFGSLFMLVGLLASAR 204

Query Match 14.6%; Score 153.5; DB 2; Length 206;
Best Local Similarity 23.4%; Pred. No. 1.9e-05;
Matches 47; Conservative 38; Mismatches 91; Indels 25; Gaps 5;

QY 17 MSPGDPFFVVSQTAVSRKREAMGVLTGCGVMWVWAGIALGLHIIIEKMAWLHTLIWV 76
: ||| | | : : : | | | | | | | | | | | | | | | | | |
Db 18 LSPGSGAINTMTTSINHGYRGAVASIAQLQTGLGIHIVLVGVGLTFLSRSLAFELKW 77
QY 77 GGGLYLCWGMQMLR--GALKKEAVSAPQVELAKSGRSFLKGLLTNLANKKAIIFYGS 134
: ||| | | | | | | | | | | | | | | | | | | | | | |
Db 78 AGAAYLILWLGIOQWRAGAILDLHTLA-----QTQSRGRLEKRAIFVNLNPKSIYFLAA 131
QY 135 VFSLEFVGDNVGTARTARIGIFALIIIVETLAWFTVVASLFPQMRRGYQRLAKWIDG----- 189
: | | | | | | | | | | | | | | | | | | | | | | | |
Db 132 LFEPQIMPQPOLAQY----LILGVT----TIVYDMVVMVTGYATLAQRIAAWIKGPKQMR 183
QY 190 ----FAGALFAGFGIHLIISR 206
: | | | | | | | | | | | | | | | | | | | | | |
Db 184 ALNKAFGSLFMLVGLLASAR 204

Search completed: May 6, 2001, 14:40:51
Job time: 247 sec
```

OM of: US-09-466-935-4 to: EST:* out_format : pfs
 Date: May 6, 2001 3:09 PM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
 -MODEL=frame*_p2n.model -DEV=xlp
 -Q=/cgn2_1/USPRO_spool/US09466935/runat_03052001_075709_14787/app_query.fasta_1.529
 -DB=EST -QFMT=fastap -SUFFIX=rst -GAPOP=12.000 -GAPEXT=4.000
 -MINMATCH=0.100 -LOOPCL=0.000 -LOPEXT=0.000 -QGAPOP=4.500
 -QGAPEXT=0.050 -XGAPOP=10.000 -XGAPEXT=0.500 -FGAPOP=6.000
 -FGAPEXT=7.000 -YGAPOP=10.000 -YGAPEXT=0.500 -DELOP=6.000
 -DELEXT=7.000 -START=1 -MATRIX=blossum62 -TRANS=human40.cdi
 -LIST=45 -DOALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0
 -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext -MINLEN=0
 -MAXLEN=2000000000 -USER=US09466935@CGN1_1_3199 -NCPU=6
 -ICPU=3 -LONGLOG -NO_XLPXY -WAIT -THREADS=1

Search information block:
 Query: US-09-466-935-4
 Query length: 206
 Database: EST:*
 Database sequences: 9623517
 Database length: 73081774
 Search time (sec): 1674.270000

score_list:	Strd	Orig	ZScore	BScore	Len	Documentation
gb_gss23:A2576789	-	134.50	308.10	4.5e-08	407	A2576789 01a12 Shot-gun genomic
gb_est89:BF614897	+	124.50	245.22	0.0001	428	BF614897 dg18a02.y1 Xenopus lae
gb_est95:BG038404	+	118.00	231.34	0.0009	443	BG038404 dg34h08.y1 Xenopus lae
gb_est52:AW863870	+	118.00	229.30	0.0011	535	AW863870 PM4-SN0012-010400-001-
gb_est52:AW863932	+	118.00	229.30	0.0011	535	AW863932 PM4-SN0012-030400-001-
gb_est96:BG116685	-	97.50	175.99	1.04	1305	BG116685 602317876F1 NIH_MGC_8
gb_gss14:AQ91044	-	94.50	178.07	0.7940	667	AQ91044 Rfc01899 Photobact
gb_gss27:R07708	+	90.50	169.10	2.51	709	R07708 736HIC0930305 Rhodobacte
gb_est79:BE872275	+	90.50	164.74	4.39	1060	BE872275 60146314F1 NIH_MGC_6
gb_est78:BE795531	+	90.50	163.54	5.12	1183	BE795531 60159251F1 NIH_MGC_7
gb_est94:BF972121	-	90.50	159.30	8.82	1749	BF972121 60224016F1 NIH_MGC_4
gb_est85:BF310123	-	90.00	157.56	11.03	1866	BF310123 60189334F1 NIH_MGC_1
gb_gss29:CI117C8	+	89.00	166.33	3.58	687	AJ226579 Clona intestinalis gen
gb_est75:BE15205	+	89.00	162.64	5.75	965	BE15205 601235912F1 NIH_MGC_44
gb_est89:BF584020	+	88.00	163.74	4.99	720	BF584020 602096258F1 NCI_CGAP_
gb_est89:BF038643	+	87.00	161.98	6.25	699	BE038643 AB03E02 AB Arabidopsis
gb_est84:BF215761	-	87.00	157.55	11.05	1052	BF215761 601881258F1 NIH_MGC_5
gb_est11:AA065225	+	86.50	163.43	5.20	556	AA065225 zms1905.s1 Strategene
gb_est83:BF132880	+	86.00	163.48	5.16	503	BF132880 601646035F1 NIH_MGC_59
gb_est24:AI778551	+	85.50	161.59	6.58	544	AI778551 EST259430 tomato succe
gb_est90:BF688728	-	85.50	153.73	18.02	1122	BF688728 602185381F1 NIH_MGC_4
gb_est97:BG251680	-	85.50	153.33	18.97	1164	BG251680 602363575F1 NIH_MGC_9
gb_est77:BE687805	+	83.50	159.88	8.19	434	BE687805 w47409.y1 Soares mam
gb_est92:BF774748	+	83.50	158.97	9.20	472	BF774748 284502 MARC 380V Bos t
gb_est73:BF301889	+	83.50	154.19	16.98	733	BE301889 601286588F1 NIH_MGC_44
gb_est88:BF530060	+	83.50	150.65	26.75	1016	BF530060 602042567F1 NCI_CGAP_
gb_est4:AA220451	+	83.00	157.84	10.64	476	AA220451 mv70007.r1 Soares mous
gb_est6:AA391827	+	83.00	156.87	12.05	473	AA391827 LD10752.5prime LD Dros
gb_est36:AI879224	+	82.50	154.18	17.01	606	AI879224 au56502.y1 Schneider H
gb_est80:BE902651	+	82.50	149.70	30.23	916	BE902651 601677332F1 NIH_MGC_21
gb_est93:BF135859	+	82.50	146.56	45.19	1223	BF135859 601782171F1 NCI_CGAP_
gb_est97:BG167246	+	82.50	143.98	62.96	1552	BG167246 602342506F1 NIH_MGC_8
gb_est74:BE455136	+	82.00	151.38	24.37	713	BE455136 HVSMEH009610f Hordum
gb_est84:BF184795	+	82.00	150.59	26.97	767	BF184795 601844538F1 NIH_MGC_54
gb_est82:BF029105	+	82.00	149.82	29.75	823	BF029105 601764618F1 NIH_MGC_53
gb_est78:BE735925	+	82.00	146.95	42.98	1072	BE735925 601305418F1 NIH_MGC_5
gb_est77:BF77883	+	82.00	142.92	72.06	1554	BF77883 602313252F1 NIH_MGC_8
gb_est14:AA950708	+	81.50	152.77	20.39	570	AA950708 LD30805.5prime LD Dros
gb_gss17:AZ203192	+	81.50	147.72	38.98	908	AZ203192 SP_0080.B1.E07 SP6E St
gb_est76:BE614696	-	81.50	140.18	102.51	1819	BE614696 601281682F1 NIH_MGC_3
gb_gss17:AZ195973	+	81.00	147.94	37.86	808	AZ195973 SP_1031.A2.A06 SP6E St

gb_est82:BF027022 - 81.00 147.60 39.56 834 ! BF027022 601671251F1 NIH_MGC
 gb_est89:BF581425 - 81.00 144.84 56.33 1075 ! BF581425 602100826F1 NCI_CG
 gb_est92:BF782441 - 81.00 140.68 96.10 1578 ! BF782441 602106149F1 NCI_CG
 gb_est12:AA817286 - 80.50 150.86 26.03 561 ! AA817286 LD22515.5prime LD D
 seq_name: gb_gss23:A2576789
 seq_documentation_block:
 LOCUS A2576789 407 bp DNA 08-DEC-2000
 DEFINITION 01a12 Shot-gun genomic library of Rhizobium strain ANU265 Rhizobium
 sp. NGR234 genomic clone 01a12, DNA sequence.
 ACCESSION A2576789
 VERSION A2576789.1 GI:11603030
 KEYWORDS GSS.
 SOURCE Rhizobium sp. NGR234.
 ORGANISM Rhizobium sp. NGR234.
 Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 Rhizobiaceae; Rhizobium.
 REFERENCE 1 (bases 1 to 407)
 AUTHORS Viprey,V., Rosenthal,A., Broughton,W.J. and Perret,X.
 TITLE Genetic snapshots of the Rhizobium species NGR234 genome
 JOURNAL Genome Biology, vol 1 (6), 0014.1-0014.7 (2000)
 COMMENT Contact: Virginie Viprey
 Laboratoire de Biologie Molculaire des Plantes Supérieures
 University of Geneva
 1 Chemin de l'Imperatrice, Chambesy/Geneva 1292, Switzerland
 Tel: +44(0)1603450000
 Fax: +44(0)1603450045
 Email: virginie.viprey@bbbsrc.ac.uk
 Class: shotgun.
 Location/Qualifiers
 1..407
 /organism="Rhizobium sp. NGR234"
 /strain="ANU265"
 /db_xref="taxon:394"
 /clone="01a12"
 /clone_lib="Shot-gun genomic library of Rhizobium strain
 ANU265"
 /note="vector: ML3; derivative strain of NGR234 cured of
 PNGR234a"
 BASE COUNT 85 a 126 c 137 g 55 t 4 others
 ORIGIN
 alignment_scores:
 Quality: 154.50 Length: 130
 Ratio: 1.797 Gaps: 2
 Percent Similarity: 66.154 Percent Identity: 30.000
 alignment_block:
 US-09-466-935-4 x A2576789/rev ..
 Align seg 1/1 to reverse of: A2576789 from: 1 to: 407
 2 LeuMetLeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSe 18
 ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 398 CTCTGGCTGTTCTTACCCTTCCTTCCTCGCATCATCATCGTC..... 357
 18 rProGlyProAspPhePheValSerGlaThrAlaValSerArgSerA 35
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 356 .CCGGCATGGACATGGTCTTCCTGCTGCGCAACGCGATAACCGCGGAC 308
 35 rGlysGluAlaMetMetGlyValLeuGlyIleThrCysGlyValMetVal 51
 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 307 GCAGTTCGGCGCTCGCGCGCGCGCGCATCAGGAGCGCGGCTGCTG 258
 52 TrpAlaGlyIleAlaLeuLeuGlyLeuHisIleLeuGlyLysMetAl 68
 : : : : : ||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
 257 CACACGCTCATGCGCGCTCGCGCGCGCGCATCAGGCTGCTGCTGCTG 208
 68 aTrpLeuHisThrLeuIleMetValGlyGlyGlyLeuThrLeuCysTrp 85
 ||| : : : : : ||||| ||||| ||||| ||||| ||||| ||||| |||||
 207 GGAGCTCTTCAAGCTGCTGCTGCTGCGCGCGCGCGCTGCTGCTGCTG 158

85 etGlyTyrGlnMetLeuArgGlyAlaLeuLeuLysGluAlaValSerAla 101
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
157 TCGGCTTCGCTGCTGCGCAGTTCATCACCATTGGCGGCTCGAGGCG 108
102 ProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuLysG1 118
||| :||| :||| :||| :||| :||| :||| :||| :||| :|||
107 GGGCGGAGGCTCTCGCGCTGGGCG.....AGCTTTCGACAGGG 70
118 yLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyr 131
| |||||:||||| ||||| ||||| :||| :||| :||| :||| :||| :|||
69 GGGCGTACAGGCTGATGAACCGAGGCTTCNCCTTC 30

seq_name: gb_est89:BF614897

seq_documentation_block: 428 bp mRNA EST 14-DEC-2000
LOCUS BF614897
DEFINITION dg18a02.y1 Xenopus laevis gastrula non normalized Xenopus laevis
cDNA clone XENOPUS_SOURCE.ID:xlnga005b04 5' similar to TR:Q9RSX2
Q9RSX2 CONSERVED HYPOTHETICAL PROTEIN. ;, mRNA sequence.
ACCESSION BF614897
VERSION BF614897.1 GI:11787968
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 428)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
, B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
WashU Xenopus EST project, 1999
Unpublished (1999)
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: <http://www.resgen.com/> please reference the id listed
below when ordering this clone: Source lab clone id - xlnga005b04
Seq primer: -40RP from Gibco
High quality sequence stop: 401.
Location/Qualifiers
1. .428
/organism="Xenopus laevis"
/db_xref="taxon:8355"
/clone="XENOPUS_SOURCE.ID:xlnga005b04"
/clone_lib="Xenopus laevis gastrula non normalized"
/tissue_type="gastrula (stages 10.5, 11.5 mixed)"
/lab_host="Top-10 F"

FEATURES

source

seq_name: gb_est95:BG038404

seq_documentation_block: 443 bp mRNA EST 24-JAN-2001
LOCUS BG038404
DEFINITION dg34h08.y1 Xenopus laevis gastrula non normalized Xenopus laevis
cDNA clone XENOPUS_SOURCE.ID:xlnga009p16 5' similar to TR:Q9RSX2
Q9RSX2 CONSERVED HYPOTHETICAL PROTEIN. ;, mRNA sequence.
ACCESSION BG038404
VERSION BG038404.1 GI:12480989
KEYWORDS EST.
SOURCE African clawed frog.
ORGANISM Xenopus laevis
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Amphibia; Batrachia; Anura; Mesobatrachia; Pipidae;
Xenopodinae; Xenopus.
1 (bases 1 to 443)
Clifton,S., Johnson,S.L., Blumberg,B., Song,J., Hillier,L., Pape,D.,
, Martin,J., Wylie,T., Underwood,K., Theising,B., Bowers,Y., Person
, B., Gibbons,M., Harvey,N., Ritter,E., Jackson,Y., McCann,R.,
Waterston,R. and Wilson,R.
WashU Xenopus EST project, 1999
Unpublished (1999)
TITLE WashU Xenopus EST project, 1999
JOURNAL Unpublished (1999)

library." 67 a 137 c 127 g 96 t 1 others
BASE COUNT
ORIGIN
alignment_scores:
Quality: 124.50 Length: 145
Ratio: 1.482 Gaps: 4
Percent Similarity: 57.931 Percent Identity: 26.897
alignment_block:
US-09-466-935-4 x BF614897 ..
Align seg 1/1 to: BF614897 from: 1 to: 428

14 ValAlaLeuMetSerProGlyProAspPhePheValSerGlnThrAl 30
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
3 ATGCTTTCGCGCTCTCCGCGCACCGCGCTGGTTTACACGCTGGCTGGCGCG 52
30 aValSerArgSerArgLysGluAlaMetMetGlyValLeuGlyIleThrC 47
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
53 CCTTTCGCGAGGGCGCGCGGCGAGGCTCATTTGGCGCTTCGGCTGCACGC 102
47 ysGlyValMetValTrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIle 63
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
103 TCGGCATCTGTCGGCATCTGCTCGCGCGCATCACCGGCTCTTTCGCGCATC 152
64 IleGluLysMetAlaTrpLeuHisThrLeuIleMetValGlyGlyLe 80
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
153 CTGCACACGAGTGGCTGGCTGGCTTTAGCGTGGTGAATATCTGGGGGCTCGC 202
80 uTyrLeuCysTrpMetGlyTyrGlnMetLeuArg.....GlyAlaLeuL 95
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
203 CTATCTGCTCTATCATGGCTGGAGACACACGCTGCAGGAGAAATGGCGCTGA 252
95 ysLysGluAlaValSerAlaProAlaProGlnValGluLeuAlaLysSer 111
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
253 AGATCATGAACGAGGCGCG.....CAAAAGGCC 284
112 GlyArgSerPheLeuLysGlyLeuLeuThrAsnLeuAlaAsnProLysAl 128
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
285 GCCCGGCTCATCGGTGAGCGCATCTCATCAATCTGCTGAACCGGAAACT 334
128 alleIleTyrPheGlySerValPheSerLeuPheValGly....AspAsnV 144
|||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
335 GTGATCTTCTTTTTCCTTTCCTCGCGAGTTCATCGCGCGCGGATGAGG 384
144 aGlyThrThrAlaArg.TrpGlyIlePheAla 154
:| ||| ||| |||:||||| |||
385 CTWTGCGGACATGGCGCATGGTCTGATCTCGGCC 417

COMMENT

Other_ESTs: dg34h08.x1
Contact: Sandy Clifton, Ph.D.
WashU Xenopus EST project, 1999
Washington University School of Medicine
4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
Tel: 314 286 1800
Fax: 314 286 1810
Email: est@watson.wustl.edu
Library constructed by Bruce Blumberg
Library normalized by Jihwan Song
DNA Sequencing by: Washington University Genome Sequencing Center
Clone distribution: Xenopus clone distribution information for
this library can be found through Research Genetics, visit their
web page at: <http://www.resgen.com/> Please reference the id listed
below when ordering this clone: Source lab clone id - xlnnga009p16
Seq primer: -40RP from Gibco
High quality sequence stop: 430.

FEATURES

```

Location/Qualifiers
1. 443
   /organism="Xenopus laevis"
   /db_xref="taxon:8355"
   /clone=XENOPUS_SOURCE_ID:xlnga009p16"
   /clone_lib="Xenopus laevis gastrula non normalized"
   /tissue_type="gastrula (stages 10.5, 11.5 mixed)"
   /lab_host="Top-10 F"
   /note="vector: pBluescript SK-; Site_1: EcoRI; Site_2:
XhoI; cDNA was prepared from 2ug of poly A+ RNA (equal
parts from stage 10.5 and stage 11.5 gastrulae).
EcoRI-XhoI cut cDNA was then ligated into Unizap-XR
(Stratagene) with EcoRI at the 5' end and XhoI at the 3'
end. SS-library phagemids were prepared by mass excision
from the original library and normalized by hybridization
to biotinylated driver (prepared from the same library by
PCR) to Cot-omega of 11. After removal of hybrids and
excess driver by streptavidin sepharose chromatography,
the ss-phagemids were made double stranded and
electroporated into Top-10 F'. Original library
construction by Bruce Blumberg (Cho et al. 1991
1111-1120). Note: This is a Xenopus Gene Collection (XGC)

```

	BASE COUNT	ORIGIN
1	0	0
2	0	0
3	0	0
4	0	0
5	0	0
6	0	0
7	0	0
8	0	0
9	0	0
10	0	0
11	0	0
12	0	0
13	0	0
14	0	0
15	0	0
16	0	0
17	0	0
18	0	0
19	0	0
20	0	0
21	0	0
22	0	0
23	0	0
24	0	0
25	0	0
26	0	0
27	0	0
28	0	0
29	0	0
30	0	0
31	0	0
32	0	0
33	0	0
34	0	0
35	0	0
36	0	0
37	0	0
38	0	0
39	0	0
40	0	0
41	0	0
42	0	0
43	0	0
44	0	0
45	0	0
46	0	0
47	0	0
48	0	0
49	0	0
50	0	0
51	0	0
52	0	0
53	0	0
54	0	0
55	0	0
56	0	0
57	0	0
58	0	0
59	0	0
60	0	0
61	0	0
62	0	0
63	0	0
64	0	0
65	0	0
66	0	0
67	0	0
68	0	0
69	0	0
70	0	0
71	0	0
72	0	0
73	0	0
74	0	0
75	0	0
76	0	0
77	0	0
78	0	0
79	0	0
80	0	0
81	0	0
82	0	0
83	0	0
84	0	0
85	0	0
86	0	0
87	0	0
88	0	0
89	0	0
90	0	0
91	0	0
92	0	0
93	0	0
94	0	0
95	0	0
96	0	0
97	0	0
98	0	0
99	0	0
100	0	0

LIBRARY.	74 a	142 c	127 g	100 t

alignment scores:

```

118.00      Length: 129
1.595      Gaps: 2
57.364      Percent Identity: 25.581

```

alignment block:

US-09-466-935-4 X BG038404

Align seq 1/1 to: BG038404 from: 1 to: 443

14 ValAlaLeuMetSerProGlyProAspPheValSerGlnThrAl 30
:
32 ATCGTTGGCCCTTCCGGGCACGGCAGATGTTTACACGTGGCGTGGCGG 81
:
30 aValSerArgSerArgLysGluAlaMetMetGlyValLeuGlyIlePheRc 47
:
82 CTTTTTCGACGGGGCGGGCGACGCATGTCGGCGCTTCGGGTGCACGC 131
:
47 ysGlyValMetValTrpAlaGlyIleAlaLeuGlyLeuHisLeuIle 63
:
132 TCGGCATCGTGGCGCATCTGCTGCCGCCATCACCGCTCTTGCGGCATC 181
:
64 IleGluLysMetalatRpleuHisThrLeuIleMetValGlyGlyLe 80
:
182 CTGCACACAGTGGCGCTGGCCCTTAGCGTGGTAAAATATCTGGGGPCGC 231
:
80 uTyrlLeuCysTrpMetGlyTyTrGlnMetLeuArg. GlyAlaLeuL 95
:
232 CTATCTGCTCACTACATGGCGTGGAAACAGCTCTCAGSAGAAATGGCGSCTGA 281
: :

[illegible]

seq_name: qb_est52:AW863870

seq_documentation_block:

LOCUS	AW863870	535 bp	mRNA	EST	22-MAY-2000
DEFINITION	PM4-SN0012-010400-001-g02			SN0012	
ACCESSION	AW863870				
VERSION	AW863870.1	GI:7997920			
KEYWORDS	EST.				
SOURCE	human.				

SOURCE	HUMAN.
ORGANISM	Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

1. (cases 1 to 535)

REFERENCE

AUTHORS

Dias Neto, E., Garcia Correa, R., Verjowski-Almeida, S., Briones, M. R., Nagai, M. A., da Silva, W. Jr., Zago, M. A., Bordin, S., Costa, F. F., Goldman, G. H., Carvalho, A. F., Matsukuma, A., Baia, G. S., Simpson, D. H., Brumstein, A., de Oliveira, P., Bucher, P., Jongeneel, C. V., O'Hare, M. J., Soares, F., Brentani, R. R., Reis, L. F., de Souza, S. J. and Simpson, A. J.

TITLE Shotgun sequencing of the human transcriptome with ORF expressed

JOURNAL
MEDLINE
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
20202663

COMMENT	Contact: Simpson A.J.G.

Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br
This sequence was derived from the FAPESP/LICR Human Cancer Genome
Project. This entry can be seen in the following URL
(<http://www.ludwig.org.br/scripts/gethtml2.pl?ti=8t2=PM4-SN0012-010>)
Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 533.

FEATURES

source
1. 535
location/quantity

```

/organism="Homo sapiens"
/db_xref="taxon.9606"
/clone_lib="SN0012"
/dev_stage="Adult"
/note="Organ: stomach_normal; Vector: puc18; Site1: SmaI;
Site2: SmaI; A mini-library was made by cloning products
derived from ORSTS PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the puc 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."
134 a 176 c 133 q 92 t
BASE COUNT

```

BASE COUNT	134 a	176 c	133 q	92 t
LOW SURVEILLANCE CONDITIONS:				

alignment_scores:

Alignment Scores:	Quality:	118.00	Length:	63
	Ratio:	2.458	Gaps:	1
	Percent Similarity:	76.190	Percent Identity:	39.683

alignment_block:
US-09-466-935-4 x AW863870/rev ..

Align seg 1/1 to reverse of: AW863870 from: 1 to: 535

5 PheLeuThrValAlaMetValHisIleValAlaLeuMetSerProGlyPr 21
|||||
235 TTTTACCGCTAGCCTTGATTCACTTGTTGGCAGTGGCGGCCGCCGCC 186

21 oAspPhePhePheValSerGlnThrAlaValSerArgSerArgLysGlu. 37
|||||
185 GGATTTCGCCGTGGTGGTGAAGCGTGACCCACGCCGCCGTGCTG 136

38 AlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAlaG1 54
|||
135 GCACATGGACGGCCCTTGGCGTGGTTCGGCGATTTCCTCATGTGGG 86

54 yIleAlaLeuLeuGlyLeuHisIleIleGluLys 66
|
85 TTTACTGTTGCTGGGTATCGGTTTGATCGTGTCCTG 49

seq_name: gb_est52:AW863932

seq_documentation_block: 535 bp mRNA EST 22-MAY-2000
LOCUS AW863932
DEFINITION PM4-SN0012-030400-001-902 SN0012 Homo sapiens cDNA, mRNA sequence.
ACCESSION AW863932
VERSION AW863932.1 GI:7997982
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 535)
Dias Neto,E., Garcia Correa,R., Verjovski-Almeida,S., Briones,M.R.,
Nagai,M.A., da Silva,W. Jr., Zago,M.A., Bordin,S., Costa,F.,
Goldman,G.H., Carvalho,A.F., Matsukuma,A., Bala,G.S., Simpson,D.H.,
Brunstein,A., deOliveira,P.S., Bucher,P., Jongeneel,C.V., O'Hare
,M.J., Soares,F., Brentani,R.R., Reis,L.F., de Souza,S.J. and
Simpson,A.J.

Shotgun sequencing of the human transcriptome with ORF expressed
sequence tags
Proc. Natl. Acad. Sci. U.S.A. 97 (7), 3491-3496 (2000)
2020663
Contact: Simpson A.J.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research
Rua Prof. Antonio Prudente 109, 4 andar, 01509-010, Sao Paulo-SP,
Brazil.
Tel: +55-11-2704922
Fax: +55-11-2707001
Email: asimpson@ludwig.org.br

This sequence was derived from the FAPESP/LICR Human Cancer Genome
project. This entry can be seen in the following URL
(http://www.ludwig.org.br/scripts/gethtml2.pl?tl=et2-pw4-SN0012-030
400-001-q02&t3=2000-04-03&t4=1)

Seq primer: puc 18 forward
High quality sequence start: 8
High quality sequence stop: 533.

FEATURES
source

1..535
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="SN0012"
/dev_stage="Adult"

/note="Organ: stomach_normal; Vector: puc18; Site_1: SmaI;
Site_2: SmaI; A mini-library was made by cloning products
derived from ORSTES PCR (U.S. Letters Patent application
No. 196,716 - Ludwig Institute for Cancer Research)
profiles into the pUC 18 vector. Reverse transcription of
tissue mRNA and cDNA amplification were performed under
low stringency conditions."

BASE COUNT 134 a 176 c 133 g 92 t

ORIGIN

alignment_scores:
Quality: 118.00 Length: 63
Ratio: 2.458 Gaps: 1
Percent Similarity: 76.190 Percent Identity: 39.683

alignment_block:
US-09-466-935-4 x AW863932/rev ..

Align seg 1/1 to reverse of: AW863932 from: 1 to: 535

5 PheLeuThrValAlaMetValHisIleValAlaLeuMetSerProGlyPr 21
|||||
235 TTTTACCGCTAGCCTTGATTCACTTGTTGGCAGTGGCGGCCGCCGCC 186

21 oAspPhePhePheValSerGlnThrAlaValSerArgSerArgLysGlu. 37
|||||
185 GGATTTCGCCGTGGTGGTGAAGCGTGACCCACGCCGCCGTGCTG 136

38 AlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAlaG1 54
|||
135 GCACATGGACGGCCCTTGGCGTGGTTCGGCGATTTCCTCATGTGGG 86

54 yIleAlaLeuLeuGlyLeuHisIleIleGluLys 66
|
85 TTTACTGTTGCTGGGTATCGGTTTGATCGTGTCCTG 49

seq_name: gb_est96:BG116685

seq_documentation_block: 1305 bp mRNA EST 30-JAN-2001
LOCUS BG116685
DEFINITION G02317878F1 NIH_MGC_88 Homo sapiens cDNA clone IMAGE:4417986 5',
mRNA sequence.

ACCESSION BG116685
VERSION BG116685.1 GI:12610191
KEYWORDS EST.
SOURCE human.

ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 1305)
NIH-MGC http://mgi.nci.nih.gov/
National Institutes of Health, Mammalian Gene Collection (MGC)
Unpublished (1999)
Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov

Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

Plate: LLAM10151 row: h column: 19
High quality sequence stop: 182.

FEATURES
source
1..1305
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone_lib="IMAGE:4417986"
/clone_lib="NIH_MGC_88"
/tissue_type="duodenal adenocarcinoma, cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: small intestine; Vector: pCMV-SPORT6;
Site_1: NotI; Site_2: SalI; Cloned unidirectionally;
oligo-dr primed. Average insert size 1.767 kb. Library
enriched for full-length clones and constructed by Life
Technologies. Note: this is a NIH_MGC Library."

BASE COUNT 457 a 463 c 347 g 38 t

ORIGIN

[illegible]

```

99 valSerAlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPh 115
100 ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
312 ...TTTCCCAATATTGAACAGGTGCAGGCGAGTAGTACAGCAAAAACGG 266
115 eLeuLysGlyLeuLeuThrAsnLeuAla.....AsnPro.... 126
116 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
265 TTGGCGGGTATTCGCGATTATCTTCGCGTCACTTGGTTAAATCCGCAVG 216
127 .....LysAlaIleIleTyrPheGlySerValPheSerLeuPheVal 140
128 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
215 TTTATTATTAGACACCATGTTGTTCTGGGCAGT.....ATT 181
141 GlyAspAsnValGlyThrThrAlaArg....TrpGlyIlePheAlaLeuL 156
142 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
180 GCGCGCCAGTTATCCTCAGAACTCAGGCCCTGTTTACTTTGGTCCGC 131
156 eIleValGluThrLeuAlaTrpPhe.....ThrValValAlaSerL 170
157 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
130 AAGCGCT...TCGTAGTTAGTTGGTTTTCCTCTGTCATTTGCTAGCGCAT 84
170 euPheAlaLeuProGlnMetArgArgGlyTyrGlnArgLeuAlaLysTrp 186
171 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
83 GGTTCCTCCCTGTTGAAGTAAGCGCGCTTCTCAACGGATT..... 43
187 IleaspGlyPhe 190
188 ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
42 ATTAATGGGTTT 31

seq_name: gb_gss27:B07708

seq_documentation_block: 709 bp DNA GSS 15-NOV-1997
LOCUS B07708
DEFINITION 736HIC0930305 Rhodobacter sphaeroides 2.4.1 genomic DNA library
Rhodobacter sphaeroides genomic clone 736HIC0930305 similar to orf
(D64002), DNA sequence.
ACCESSION B07708
VERSION B07708.1 GI:2055500
KEYWORDS GSS.
SOURCE Rhodobacter sphaeroides.
ORGANISM Rhodobacter sphaeroides;
Bacteria; Proteobacteria; alpha subdivision; Rhodobacter group;
Rhodobacter.
REFERENCE 1 (bases 1 to 709)
AUTHORS Choudhary,M., Mackenzie,C., Nereng,K., Sodergren,E., Weinstock,G.M.
and Kaplan,S.
TITLE Low-resolution sequencing of Rhodobacter sphaeroides 2.4.1T:
chromosome II is a true chromosome
JOURNAL Microbiology 143, 3085-3099 (1997)
MEDLINE 98015398
COMMENT Contact: Mackenzie, C.
Department of Microbiology and Molecular Genetics
The University of Texas Health Science Center
6431 Fannin St., Houston, TX 77006, USA
Tel: (713) 794-1742
Fax: (713) 794-1782
Email: mackenz@utmsg.med.utmc.edu
Seq primer: pBluescript SK (-) T3
Class: shotgun.
Location/Qualifiers
1..709
/organism="Rhodobacter sphaeroides"
/strain="2.4.1T"
/db_xref="taxon:1063"
/clone="736HIC0930305"
/clone_lib="Rhodobacter sphaeroides 2.4.1 genomic DNA
library"
/lab_host="E. coli SL7-1"
/note="Vector: pLA2917; Genomic DNA from Rhodobacter
sphaeroides was prepared and partially digested with
Sau3AI. Size selected (20kb) fragments were subcloned into
the BglII site of the cosmid vector pLA2917 (Allen, L. N.
and R. S. Hanson, 1985. Journal of Bacteriology,
161:955-962. The library was then ordered around

```

```

chromosome II of Rhodobacter sphaeroides (Choudhary et
al., 1994. Journal of Bacteriology, 176:7694-7702). The
cosmids were then digested with a variety of restriction
enzymes (BamHI, EcoRI, BglII, PstI, EcoRV, NotI and
DNaseI) and restriction fragments subcloned into the
respective multiple cloning site sites of pBluescript SK
(-). Note BglII fragments were subcloned into the
pBluescript BamHI site. DNaseI fragments were subcloned
into the EcoRV site. All subclones were transformed into
E. coli XL1Blue MRF+. All fragments were then sequenced
and the sequences where possible were assembled using the
CGC program GELASSEMBLE."

BASE COUNT 125 a 200 c 210 g 166 t 8 others
ORIGIN

alignment_scores:
Quality: 90.50 Length: 79
Ratio: 1.926 Gaps: 1
Percent similarity: 59.494 Percent identity: 29.114

alignment_block:
US-09-466-935-4 x B07708 ..

Align seg 1/1 to: B07708 from: 1 to: 709

4 LeuPheLeuThrValAlaMetValHisIleValAlaLeuMetSerProG1 20
|||||: |||: ||||| : ||||| : ||||| : ||||| : ||||| : |||||
467 CTCTTCATGTTGCTGCTCGCGTG...ATCGTTCTGACNATCAGCGCGG 513

20 yProAspPhePheValSerGlnThrAlaValSerArgSerArgLysG 37
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
514 CCGCGATCTTTTCCTGATCTCGGCGCGGATCTCTCAAGTGGGCGCG 563

37 luAlaMetMetGlyValLeuGlyIleThrCysGlyValMetValTrpAla 53
|||||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
564 CCGCCTGTTCACCGCTTTCGGCTTCTTCGCGGCGTTCATCAAGTG 613

54 GlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMetAlaTrpLe 70
|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
614 CGCTGCTTGGCTTGGATGCTGCCNACGCTTGTCTGGCAGCCCGCGCT 663

70 uHisThrLeuIleMetValGlyGlyGlyLeuTyrLeu 82
|||: |||||: |||||: |||||: |||||: |||||: |||||: |||||
664 TTTCGACNCTCGCTGCTTATGGGGTGGTGCATATCTT 700

seq_name: gb_est79:BB872275

seq_documentation_block: 1060 bp mRNA EST 20-OCT-2000
LOCUS BB872275
DEFINITION 601446314F1 NIH_MGC_65 Homo sapiens cDNA clone IMAGE:3850491 5',
mRNA sequence.
ACCESSION BB872275
VERSION BB872275.1 GI:10321051
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1060)
AUTHORS NIH-MGC http://mhc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov

```

```
Plate: LAM9569 row: k column: 04
High quality sequence stop: 468.
FEATURES
    source
        1..1060
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_lib="IMAGE:3850491"
            /tissue_type="NIH_MGC_65"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: colon; Vector: pCMV-SPORT6; Site_1: NotI;
            Site_2: SalI; Cloned unidirectionally. Primer: Oligo dr.
            Average insert size 1.8 kb. Library constructed by Life
            Technologies."
BASE COUNT    309 a    378 c    283 g    90 t
ORIGIN
alignment_scores:
    Quality: 90.50    Length: 279
    Ratio: 0.838    Gaps: 13
    Percent Similarity: 38.710    Percent Identity: 19.713
alignment_block:
US-09-466-935-4 x BE872275/rev ..
Align seg 1/1 to reverse of: BE872275 from: 1 to: 1060
13 lleValAlaLeuMetSerProGlyProAspPhePheValSerGlnTh 29
:::|||||:::|||||:::|||||:::
903 CTGGTCGGGCTCTCTCTGGGTTTCCTCGC.....GCTGTTCT 866
29 rAlaValSerArgSerArgLysGluAlaMetMetGly.....41
|||||:::|||||:::|||||:::
865 GCCCGTCTCGGCCTTCGCTCGATGCTCTCCGGGGCTGGTGTCTCTAG 816
42 .....ValLeuGlyIleThrCysGlyVal 49
|||||:::|||||:::|||||:::
815 GCTGGAGTGCTATGGTTCCTGTTGCTCTGCTGTTCTGTTGCGGTGT 766
50 MetVal.....51
765 TGTGTGGGTTCTGTGTTGTGGGTCGTGTTTCCGTGTTGGGTTGGCCGTG 716
52 TrpAlaGlyIleAlaLeuLeuGlyLeuHisLeuIleIleGluLysMet. 67
|||||:::|||||:::|||||:::
715 TTGGCGTGGGTTGCT...CTGTGTTGCGGTTTGTGTCCTAGTTGG 669
68 .....AlaTrpLeuHisThr 72
|||||:::|||||:::
668 TACGATGTTGGGTGTCGTGCGCTTCTTGGTGGTGTGTGTGGAGCGTTTCT 619
73 LeuIleMetValGlyGlyLeuTyLeuTyLeuCys.....TrpMe 85
|||||:::|||||:::|||||:::
618 GTGCTGTCTCTGGGTGGTGGTCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 569
85 t.....GlyTyLeuMetL 90
|||||:::|||||:::
568 GTGTCAAGTCTCGAGTTTGTAGGAGTGACTGCTCTCGGTGGTTTC.....G 525
90 euArgGlyAlaLeuLys.....LysGluAlaValSer 100
|||||:::|||||:::
524 CTTCCGGGTGCTCTCCCGGTGTCGCTGCTGCTGCTGCTGCTGCTGCTGCTC 475
101 AlaProAlaProGln.....105
|||||:::|||||:::
474 GCGCAGCGCCGAGACCCAGCAGCGTCTGCCAGCGAGAGCGGTAGACAC 425
106 .....ValGluLeuAlaLysSerGlyArgSerPheLeuLysGlyL 119
|||||:::|||||:::
424 GCGCGTGTACTGTGGGCGAGGCTCAGGACCCCGCTTCCAGTCCACAC 375
119 euLeuThrAsnLeuAlaAsnProLysAlaIleIleTyLeuPheGlySerVal 135
```

```
|||||:::|||||:::|||||:::
374 TCCTGTCCGGCTTGGTGTCTCCAGGTCTGTACAGCTGTCTGGGCCCTTC 325
136 PheSerLeuPheVal.....G1 141
|||||:::|||||:::
324 TTCTTCTTGGCTTTCAGATGAGGACATCCCGCTCCGGGAATGGGG 275
141 yAspAsnValGlyThrAlaArgTrpGlyIlePheAlaLeuIleLev 158
:::|||||:::|||||:::|||||:::
274 CATCTCTCCAGCAGCTGCCAGAAATGTCGGTGTGCGCCACGGTGTCTGA 225
158 alGluThr.....Leu 161
:::|||||:::
224 CGCGAGGTACTCCACAGCAGCTGCTGCAGCTCCACGCTCGCTGCTCTG 175
162 AlaTrpPheThrValValAlaSerLeuPheAlaLeuProGlnMetArgAr 178
:::|||||:::
174 AGCTGGCTGTCTGCTG.....160
178 gGlyTyLeuGlnArgLeuAlaLysTrpIleAspGlyPhe 190
|||||:::|||||:::
159 .....CGCAGCAGCTCTCGGATGCTGGGCTTC 133
seq_name: gb_est78:BE795531
seq_documentation_block:
LOCUS BE795531 1183 bp mRNA EST 20-SEP-2000
DEFINITION 601592515F1 NIH_MGC_7 Homo sapiens cDNA clone IMAGE:3946562 5',
mRNA sequence.
ACCESSION BE795531
VERSION BE795531.1 GI:10216729
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1183)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/.
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: DCTD/DTp
cDNA Library Preparation: Ling Hong/Rubin Laboratory
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at: Image.llnl.gov
Plate: LLCM806 row: n column: 03
High quality sequence stop: 791.
FEATURES
    Location/Qualifiers
        1..1183
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone_image="IMAGE:3946562"
            /clone_lib="NIH_MGC_7"
            /tissue_type="small cell carcinoma"
            /cell_line="MGC3"
            /lab_host="DH10B (phage-resistant)"
            /note="Organ: lung; Vector: pORF7; Site_1: XhoI; Site_2:
            EcoRI; cDNA made by oligo-dT priming. Directionally
            cloned into EcoRI/XhoI sites using the following 5',
            adaptor: GGCACGAG(G). Size-selected >500bp for average
            insert size 1.8kb. Library constructed by Ling Hong in
            the laboratory of Gerald M. Rubin (University of
            California, Berkeley) using ZAP-cDNA synthesis kit
            (Stratagene) and Superscript II RT (Life Technologies)."
```

BASE COUNT 264 a 319 c 344 g 254 t

ORIGIN alignment_scores:

```
Quality: 90.50 Length: 219
Ratio: 0.943 Gaps: 10
Percent Similarity: 43.836 Percent Identity: 23.744

alignment_block:
US-09-466-935-4 x BE795531 ..

Align seg 1/1 to: BE795531 from: 1 to: 1183

16 LeuMetSerProGlyProAspPhePheValSerGlnThrAlaValSe 32
||||| ||| :|| :||| :||| :||| :||| :||| :|||
411 CTGATGCTGAGGGGAGTACTTTTGAACCTTCTGACCGAGTCGAGT 460
32 rAqSerArgLysGluAlaMetMetGlyVal.....L 43
:||||| :||| :||| :||| :||| :||| :||| :|||
461 TCAGGCTGGACCTAGCAGTGTGAGAGCCACTGAGCGAGAGTCTCT 510
43 euGlyIleThrCysGlyValMetValTrpAlaGlyIleAlaLeuGly 59
:||||| :||| :||| :||| :||| :||| :||| :|||
511 TTGGGAGTCACTGTGGT.....TTAGGT 533
60 LeuHisLeuIleIleGluLysMetAlaTrpLeuHisThrLeuIleMetVa 76
:||||| :||| :||| :||| :||| :||| :||| :|||
534 TTTTCTTCCCATC.....CCACTTAAGGTGAT 562
76 lGlyGlyGly.....LeuTyLeuCysT 84
:||||| :||| :||| :||| :||| :||| :||| :|||
563 GGGGCAAGGAAGAACTACAGTCCCTCCCTGCAATATATTTGTATGCT 612
84 rPMetGlyTrpGlnMetLeuArgGlyAlaLeuLysGluAlaValSer 100
:||||| :||| :||| :||| :||| :||| :||| :|||
613 GGGTGGCTGGCTGAGCTCAGAGGCTCTTAGAGAGGACACTACATCC 662
101 AlaProAlaProGlnVal.....G 107
:||||| :||| :||| :||| :||| :||| :||| :|||
663 CTTCCACCACTGGATGCCATTTCTGAGTCAAGTCACTGAAGTGAGA 712
107 uLeuAlaLysSerGlyArg..... 113
:||||| :||| :||| :||| :||| :||| :||| :|||
713 GTGTGCTCCCAAGGAGGGCTCTCTCCATCAGGATGGGTACTGTGGG 762
114 .....SerPheLeuLysGlyLeuLeuThr 121
:||||| :||| :||| :||| :||| :||| :||| :|||
763 GGAACAAATAGTACAGGCTATTGGTTCCCTTTTGAAGGTGCTGCTGAA 812
122 AsnLeuAlaAsn.....ProLysAlaIleIleTy 131
:||||| :||| :||| :||| :||| :||| :||| :|||
813 GCACTAAAGATGATGATCCAGGGGCTCCTCTCCATCAGGATGGGTGCTG... 860
131 rPheGlySerValPheSerLeuPheValGlyAspAsnValGlyThr.... 146
:||||| :||| :||| :||| :||| :||| :||| :|||
861 ....GCTTCTACTATTTCGCTCTCTAGAGGCCCCCAAGGTCCGACGAGCT 906
147 ..ThrAlaArgTTPGlyIlePheAlaLeuIleIleValGluThrLeuAla 162
:||||| :||| :||| :||| :||| :||| :||| :|||
907 GGACAGTACTGCCCCCCTCTCTCAGAGCAAGGTCCCTTGTGTGCT 956
163 TrpPheThrValValAla.....SerLeuPheAlaLeuProGlnMetar 177
:||||| :||| :||| :||| :||| :||| :||| :|||
957 AGATGGCGGTAGCGAGGATAGGATACAAAGGTTGCCACGCGCGGA 1006
177 qArgGly 179
:||||| :||| :||| :||| :||| :||| :||| :|||
1007 AAGGGGT 1013
```

seq_name: gb_est94:BF972121

seq_documentation_block:

LOCUS BF972121 1749 bp mRNA

DEFINITION 602240161F1 NIH_MGC_46 Homo sapiens cDNA clone IMAGE:4328918 5',

ACCESSION BF972121

VERSION BF972121.1 GI:12339336

KEYWORDS EST. human.
SOURCE Homo sapiens
ORGANISM Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 1749)
AUTHORS NIH-MGC http://mgc.nci.nih.gov/
TITLE National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL Unpublished (1999)
COMMENT Contact: Robert Strausberg, Ph.D.
Tel: (301) 496-1550
Email: Robert_Strausberg@nih.gov
Tissue Procurement: ATCC
cDNA Library Preparation: Ling Hong/Rubin Laboratory
DNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
cDNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLCMI189 row: i column: 15
High quality sequence stop: 329.

FEATURES

Location/Qualifiers
1..1749
/organism="Homo sapiens"
/db_xref="taxon:9606"
/clone="IMAGE:4328918"
/tissue_type="leiomyosarcoma cell line"
/lab_host="DH10B (phage-resistant)"
/note="Organ: uterus; Vector: pOTB7; Site 1: XhoI; Site 2:
EcoRI; cDNA made by oligo-dT priming. Directionally cloned
into EcoRI/XhoI sites using the following 5' adaptor:
GGCAGCAG(C). Size-selected >500bp for average insert size
1.8kb. Library constructed by Ling Hong in the laboratory
of Gerald M. Rubin (University of California, Berkeley)
using ZAP-cDNA synthesis kit (Stratagene) and Superscript
II RT (Life Technologies). Note: this is a NIH_MGC
Library."
BASE COUNT 623 a 603 c 373 g 150 t
ORIGIN

alignment_scores:

Quality: 90.50 Length: 172
Ratio: 1.104 Gaps: 12
Percent Similarity: 47.674 Percent Identity: 24.419

alignment_block:

US-09-466-935-4 x BF972121/rev ..

Align seg 1/1 to reverse of: BF972121 from: 1 to: 1749

41 GlyValLeuGly...IleThrCysGlyValMetValTrpAlaGlyIleAl 56
||||| :||| :||| :||| :||| :||| :||| :|||
855 GGTGTCTGGCGGGGTGGGGTGGCGCTTTGTCTATGACGCTCTTC 806
56 aLeuLeuGlyLeuHis.....LeuIleIleGluLysMetAlaTrpLeuH 71
:||||| :||| :||| :||| :||| :||| :||| :|||
805 GTCGTGGGTGTGTCTCTCTAGTTGGTGT.....TGTGGGTCT 765
71 isThrLeuIleMetValGlyGlyLeuTyLeuCysTrpMetGly... 86
:||||| :||| :||| :||| :||| :||| :||| :|||
764 TTTCTGTGTAGTCGTGGGGGACCCCTATGG...TGTGGAGTGGGCG 718
87TyrGlnMetLeuArgGlyAlaLeuLysLysGluAlaValSe 100
:||||| :||| :||| :||| :||| :||| :||| :|||
717 TTTGTTGTATACTATTATGTCGTGGGTCTCTCTCGCGG...GCGGTAC 671
100 rAlaProAlaProGlnValGluLeuAlaLysSerGlyArgSerPheLeuL 117
:||||| :||| :||| :||| :||| :||| :||| :|||
670 AGCGTCGCCCTGTGTC..... 652
117 ysglyLeuLeuThrAsnLeuAlaAsnProLysAlaIleIleTyPheGly 133

Clark,V.H. and Bird,A.
Nonmethylated transposable elements and methylated genes in a
chordate genome
Science 283 (5405), 1164-1167 (1999)
99148102
REFERENCE
3 (bases 1 to 687)
Simmen,M.W. and Bird,A.
Sequence analysis of transposable elements in the sea squirt, ciona
intestinalis
Mol..Biol. Evol. 17 (11), 1685-1694 (2000)
20523971
4 (bases 1 to 687)
Clark,V.H., Leitgeb,S., Charlton,J., Tweedie,S., Simmen,M.W. and
Bird,A.P.
Direct Submission
Submitted (28-FEB-1998) ICMB, University of Edinburgh, King's
Buildings, Mayfield Rd, Edinburgh EH9 3JR, UK. E-mail contact:
VCLARK@serv0.bio.ed.ac.uk
Vector: pBluescript KS;
FEATURES
source
1..687
/organism="Ciona intestinalis"
/db_xref="taxon:7719"
/clone="17C8"
/dev_stage="adult"BASE COUNT 132 a 143 c 188 g 221 t 3 others
ORIGIN

alignment_scores:
Quality: 89.00 Length: 240
Ratio: 0.881 Gaps: 13
Percent Similarity: 42.083 Percent Identity: 24.583

alignment_block:
US-09-466-935-4 x CII7C8 ..

Align seg 1/1 to: CII7C8 from: 1 to: 687

6 LeuthrValAlaMetValHisIleValAlaLeuMetSerProGlyProAs 22
||||||| :||| ||||| :|||
53 TTAACCGTATTATTAGCACTTTCTCATGCTGATGGCAGCTGTGCTA 102
22 pPhephepheValserGlnThrAlaValSerArgSerArgLysGluaLam 39
:::||| |||||::||| ||| ::|||::|||
103 TTGGTTCAGCGTGCTGAATCACTCTGGTCAGTATCTTATCCCGCTCGC 152
39 etMetGlyValLeuclyIleThrCysGlyValMetValTrpAlaGlyIle 55
:|||||||::||| ||| ::|||::|||
153 TTATGGGAGTGATGCCACACCGTGT.....ATTGTTGCCTCCCTG 193
56 AlaleuLeuGlyLeuHisLeuIleLeuGluLysMetAlaTrpLeu..... 70
:|||||
194 ACGCTG.....GCTTCCGTTGGATGCGAGA 219
71HisthrLeuIleMetValG 77
220 GGATCAGTTCCCGCAATGCGGGTATGTTGGAAGCTTTGCCATGCTGG 269
77 lyGlyGlyLeu..... 80
270 GTGGTGGCCCTCGGGCCGATGCTCTCCGAACCTGCTTGATGATGCGGC 319
81 TyrLeuCystTrpMetGlyTyrgInMetLeuArgGlyAla..... 93
::: ||| ||| ||| |||||
320 TGCCAATGTACCATAAAATCGTCGCTTTAGTCGGTGTGTAUATTCATT 369
94LeuLysGluAlaValSerAlaProAlaProGlnv 106
370 TTTAAATTTTGTATGTTAAAANTCCCCGTCATACCCAGTATCGGAAIT 419
106 alGluLeuAlaLysSerGlyArg.....SerPheLeuLysGly... 118
||| ||||| ||| :||| |||||

420	CTGAGACGGATATATCTGGCGGTATTGTTCACATTTTAAGAAAGCGAA	469
119	...LeuLeuThrAsnLeu.....AlaAsnProLysAlaIleIleTyr.P 132	
470	GACCTGAGACAGAACCTTTAATCGAATGGCATACGCTCAATGCTGATT	519
132	heGly.....SerValPheSerLeuPheValGlyAspAsnVal 144	
520	TCGGATTGGATTGTTTCAATCATTTCAACGTTTGGAGG..... 560	
145	GlyThrThrAlaargtrpglyllePheAlaLeuIleValGluThrLe 161	
561CTGGGGGATCCAT..... 575	
161	uAlatrppheThrValValAlaSerLeuPheAlaLeuProGlnMetArgA 178	
576TTTGTATGAACGCTTTCGCCGGCANGAAAA 606	
178	rgGlyTyrGlnArgLeuAlaLysTrpIleAspGlyPheAlaGlyAlaLeu 194	
607	ACGGTTGG.....GGAGTGGTGGGTGTATG 632	
195	PheAlaGlyPheGlylle 200	
633	TTTCTGGGGCGGGGTATA 650	
seq_name: gb_est75:BE515205		
seq_documentation_block: 965 bp mRNA EST 07-AUG-2000		
LOCUS	BE515205	
DEFINITION	601235912F1 NIH_MGC_44 Homo sapiens cDNA clone IMAGE:3608091 5',	
	mRNA sequence.	
ACCESSION	BE515205	
VERSION	BE515205.1 GI:9722420	
KEYWORDS	EST.	
SOURCE	human.	
ORGANISM	Homo sapiens	
	Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi	
	Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.	
REFERENCE	1 (bases 1 to 965)	
AUTHORS	NIH-MGC http://mgc.nci.nih.gov/ .	
TITLE	National Institutes of Health, Mammalian Gene Collection (MGC)	
JOURNAL	Unpublished (1999)	
COMMENT	Contact: Robert Strausberg, Ph.D. Tel: (301) 496-1550 Email: Robert.Strausberg@nih.gov Tissue Procurement: ATCC CDNA Library Preparation: Ling Hong/Rubin Laboratory CDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL) DNA Sequencing by: Incyte Genomics, Inc. Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov Plate: LLCM260 row: g column: 04 High quality sequence stop: 655.	
FEATURES	Location/Qualifiers	
source	1..965	
	/organism="Homo sapiens"	
	/db_xref="taxon:9606"	
	/clone="IMAGE:3608091"	
	/clone_lib="NIH_MGC_44"	
	/tissue_type="endometrium, adenocarcinoma cell line"	
	/lab_host="DH10B (phage-resistant)"	
	note="Organ: uterus; Vector: pOTB7; Site:1: XhoI; Site	
	EcoRI; cDNA made by oligo-dT priming. Directionally	
	cloned into EcoRI/XhoI sites using the following 5'	
	adaptor: GGCACAG(G). Library constructed by Ling Hong	
	in the laboratory of Gerald M. Rubin (University of	
	California, Berkeley) using ZAP-cDNA synthesis kit	
	(Stratagene) and superscript II RT (Life Technologies)	
BASE COUNT	199 a 292 c 259 g 215 t	
ORIGIN		

